

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
3 May 2001 (03.05.2001)

PCT

(10) International Publication Number
WO 01/30843 A1

(51) International Patent Classification⁷: **C07K 14/47**,
14/72, C12N 15/62, 15/86, A61K 38/17, C07K 14/705

BEERLI, Roger; 4364 Robbins St., San Diego, CA 92122
(US).

(21) International Application Number: PCT/EP00/10430

(74) Agent: BECKER, Konrad; Novartis AG, Corporate
Intellectual Property, Patent & Trademark Department,
CH-4002 Basel (CH).

(22) International Filing Date: 23 October 2000 (23.10.2000)

(81) Designated States (national): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,
TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/433,042 25 October 1999 (25.10.1999) US
09/586,625 2 June 2000 (02.06.2000) US

(84) Designated States (regional): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

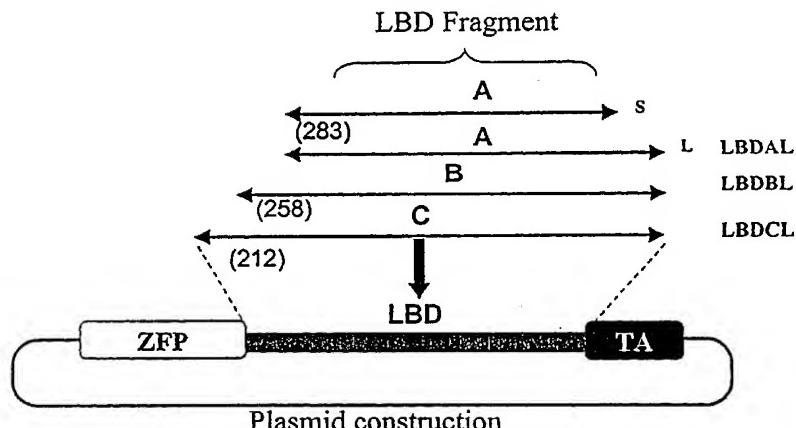
(71) Applicants: NOVARTIS AG [CH/CH]; Schwarzwaldallee
215, CH-4058 Basle (CH). THE SCRIPPS
RESEARCH INSTITUTE [US/US]; 10550 North Torrey
Pines Road, La Jolla, CA 92037 (US).

Published:

— With international search report.

[Continued on next page]

(54) Title: LIGAND ACTIVATED TRANSCRIPTIONAL REGULATOR PROTEINS



ZFP	Binding Sequence	Chimerics Constructed
C7	GCG TGG GCG	AS, BS, CS, AL, BL, CL
E2C	GGG GCC GGA g	AS, BS
2C7	GCG TGG GCG GCG TGG GCG	AS, BS, CS

WO 01/30843 A1

(57) Abstract: Fusion proteins for use as ligand-dependent transcriptional are provided. The fusion proteins include a nucleotide binding domain operatively linked to a ligand-binding domain. They also can include a transcription regulating domain. The nucleotide binding domain is a zinc-finger peptide that binds to a targeted contiguous nucleotide sequence of from 3 to 18 nucleotides are provided. The fusion proteins are used for gene therapy. Also provided are polynucleotides encoding the fusion proteins, expression vectors, and transfected cells.



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

LIGAND ACTIVATED TRANSCRIPTIONAL REGULATOR PROTEINS

Work described herein was supported by National Institutes of Health NIH Contract No. GM53910. The United States Government has certain rights in such subject matter.

RELATED APPLICATIONS

5 This application is a continuation-in-part of U.S. application Serial No. 09/433,042, filed October 25, 1999, to Carlos F. Barbas III, Michael Kadan, and Roger R. Beerli, entitled "Recombinant Ligand Activated Transcriptional Regulator Polypeptides." U.S. application Serial No. 09/433,042 is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

10 The field of this invention is the regulation of gene expression. In particular, ligand-activated fusion proteins (also referred to herein as chimeric regulators) and the use thereof for regulation of gene expression are provided. The fusion polypeptides contain a DNA binding domain containing one or a plurality of zinc finger polypeptide domains and a ligand binding domain (LBD) derived from an intracellular receptor.

BACKGROUND OF THE INVENTION

15 Intracellular receptors are a superfamily of related proteins that mediate the nuclear effects of a variety of hormones and effector molecules, include steroid hormones, thyroid hormones and vitamins A and D. Members of this family of intracellular receptors are prototypical ligand activated transcription factors. These receptors contain two primary functional domains: a DNA binding domain (DBD) that contains about sixty-six amino acids and a ligand-binding domain (LBD) located in the carboxyl-terminal half of the receptor that has about 300 amino acids. The receptors are inactive in the absence of hormone (ligand) by virtue of association with inactivating factors, such as heat shock proteins. Upon ligand binding, the receptors dissociate from the inactivating complex and dimerize, which renders them able to bind to DNA and modulate transcription.

For example, for the steroid receptors, binding of a steroid hormone to its receptor results in receptor protein homodimerization and subsequent binding to the "steroid response element" (SRE) DNA sequence in nuclear DNA. Conformational changes in the receptor associated with ligand binding results in the recruitment of other transcriptional regulatory proteins, called co-activators, that regulate the transcription from promoters adjacent to the SRE binding sites.

Modified steroid hormone receptors have been developed for use for regulated expression of transgenes (see, e.g., U.S. Patent No. 5,874,534 and published International PCT application No. WO 98/18925, which is based on U.S. provisional application Serial No. 60/029,964) by modifying the ligand specificity of the LBD. In addition, the DNA binding domain of the receptor has been replaced with a non-mammalian DNA binding domain selected from yeast GAL4 DBD, a viral DBD and an insect DBD binding domain to provide for regulated expression of a co-administered gene containing a region recognized by the non-mammalian DBD. These constructs, however, have several drawbacks. The non-mammalian DBD is potentially immunogenic and the array of sequences recognized by these DBD is limited, thereby severely restricting gene targets.

Therefore, there remains a need for more versatile gene regulators. It is an object herein to provide polypeptides that function as versatile regulators of gene expression.

SUMMARY OF THE INVENTION

Polypeptides that function as ligand activated transcriptional regulators and nucleic acid molecules encoding such polypeptides are provided. The polypeptides are fusion proteins that are ligand activated transcriptional regulator that can be targeted to any desired endogenous or exogenous gene. Variants of the fusion protein can be designed to have different selectivity and sensitivity for endogenous and exogenous ligands.

Nucleic acid molecules encoding the fusion proteins, expression vectors containing the nucleic acids and cells containing the expression vectors are provided. The fusion protein or nucleic acids, particularly vectors, that encode the fusion protein can be introduced into a cell and, when expressed in the cell, regulate gene expression in a ligand-dependent manner.

5 Fusion proteins

The fusion proteins provided herein contain a ligand binding domain (designated herein LBD) from an intracellular receptor, preferably a LBD 10 that has modified ligand specificity compared to the native intracellular receptor from which the LBD originates, and a nucleic acid binding domain (designated herein DBD) that can be tailored for any desired specificity. The fusion proteins may also include a transcriptional regulating domain (designated herein TRD), particularly a repressor or 15 activator domain. The domains are operatively linked whereby the resulting fusion protein functions as a ligand-regulated targeted transcription factor.

When delivered to the nucleus of a cell, the domains, which are 20 operatively linked, together act to modulate the expression of a targeted gene, which may be a native gene in a cell or a gene that also is delivered to a cell. Hence the targeted gene can be an endogenous cellular gene or an exogenously supplied recombinant polynucleotide construct. The fusion protein may also include a transcriptional regulating domain that is selected to activate, enhance or suppress transcription of a targeted gene.

25 In one embodiment, the fusion protein is constructed from components highly similar to human proteins, preferably components that are about 80% more preferably about 85%, most preferably at least about 90% identical in amino acid sequence to the corresponding human domain. In another embodiment, the fusion protein binds to a naturally occurring gene and modulates the transcription of the naturally occurring 30 gene in a ligand-dependent way. In another embodiment, the fusion

protein binds to an exogenously supplied recombinant construct and modulates the transcription of the exogenously supplied recombinant construct in a ligand-dependent way.

In a preferred embodiment, the isolated recombinant fusion protein forms a dimer when bound to a polynucleotide. The dimer can be a homodimer or a heterodimer. In one embodiment, the dimer includes at least one DNA binding domain, at least one, preferably two, ligand binding domains and at least one transcription modulating domain. In heterodimers, the dimer can include two different DNA binding domains, two different ligand binding domains or two different transcription modulating domains. One exemplary heterodimer includes at least three zinc finger modular units, two different ligand binding sites and a transcription modulating domain.

Exemplary fusion proteins containing zinc fingers and LBD that are non-responsive to estrogen, and that are induced by synthetic non-steroidal drugs that are routinely used for clinical treatments are described; these regulators provide ligand-dependent gene activation. Exemplary fusion proteins comprise the sequence of amino acids encoded by the open reading frame set forth in each of SEQ ID Nos. 1-18.

The fusion proteins can be used in plant species as well as animals. Transgenic plants resistant to particular bacterial or viral pathogens can be produced.

Ligan Binding Domain (LBD)

The LBD is derived from an intracellular receptor, particularly a steroid hormone receptor. The receptors from which the LBD is derived include, but is not limited to, glucocorticoid receptors, mineralocorticoid receptors, thyroid hormone receptors, retinoic acid receptors, retinoid X receptors, Vitamin D receptors, COUP-TF receptors, ecdysone receptors, Nurr-1 receptors, orphan receptors and variants thereof. Receptors of these types include, but are not limited to, estrogen receptors, progesterone receptors, glucocorticoid- α receptors, glucocorticoid- β

receptors, androgen receptors and thyroid hormone receptors. LBDs preferably are modified to alter ligand specificity so that they preferentially bind to an exogenous ligand, such as a drug, compared to an endogenous ligand.

5 When intended for human gene therapy, the ligand binding domain preferably retain sufficient identity, typically at least about 90% sequence identity to a human ligand binding domain, to avoid substantial immunological response. A single amino acid change in the LBD can dramatically alter performance of the protein.

10 The LBD is preferably modified so that it does not bind to the endogenous ligand for the receptor from which the LBD is derived, but to a selected ligand to permit fine tuned regulation of targeted genes. Hence, in certain embodiments, the ligand-binding domain has been modified to change its ligand selectivity compared to its selective in the native receptor. Preferably the modified ligand-binding domain is not substantially activated by endogenous ligands. Any method for altering ligand specificity, including systematic sequence alteration and testing for specificity, and selection protocols (see, e.g., U.S. Patent No. 5,874,534 and Wang *et al.* (1994) *Proc. Natl. Acad. Sci. U.S.A.* 91:8180-8184) can
15 be used.
20

Nucleic acid binding domain (DBD)

To achieve targeted and specific transcriptional regulation the DBD includes at least one zinc finger modular unit and is engineered to bind to targeted genes. The zinc finger nucleic acid binding domain contains at
25 least two zinc finger modules that bind to selected sequences of nucleotides. Any zinc finger or modular portions thereof can be used. The DBD replaces or supplements the naturally-occurring zinc finger domain in the receptor from which the ligand binding domain is derived.

The nucleic acid binding domain (DBD) includes at least one, preferably at least two, modular units of a zinc finger nucleic acid binding polypeptide, each modular unit specifically recognizing a three nucleotide

sequence of bases. The resulting DBD binds to a contiguous sequence of nucleotides of from 3 to about 18 nucleotides.

As noted, the DBD contains modular zinc-finger units, where each unit is specific for a trinucleotide. Modular zinc protein units can be combined so that the resulting domain specifically binds to any targeted sequence, generally DNA, such that upon binding of the fusion protein to the targeted sequence transcription of the targeted gene is modulated.

The zinc finger-nucleotide binding portion of the fusion protein can be derived or produced from a wild type zinc finger protein by truncation or expansion, or as a variant of a wild type-derived polypeptide by a process of site directed mutagenesis, or by combination of a variety of modular units or by a combination of procedures.

Cys₂His₂ (C2H2) type zinc finger proteins are exemplary of the zinc fingers that can replace the naturally occurring DNA binding domain in an intracellular receptor, such as the C4-C4 type domain in a steroid receptor, to form a functional ligand-responsive transcription factor fusion protein. By virtue of the zinc finger, the resulting fusion protein exhibits altered DNA binding specificity compared to the unmodified intracellular receptor.

The optimal portion of the ligand binding domain (LBD) of the receptor to use, the zinc finger array and extent thereof and the stoichiometry and orientation of DNA binding can be empirically determined as exemplified herein for a steroid receptor.

In preferred embodiments the zinc-finger portion of the fusion protein binds to a nucleotide sequence of the formula (GNN)_n, where G is guanidine, N is any nucleotide and n is an integer from 1 to 6, and typically n is 3 to 6. Preferably, the zinc-finger modular unit is derived from C2H2 zinc-finger peptide. More preferably, the zinc-finger peptide is a C2H2 zinc-finger peptide has at least 90% sequence identity to a human zinc-finger peptide.

Transcription Regulating Domain (TRD)

The fusion proteins also can include transcription regulating domains. In preferred embodiments, the transcription regulating domain includes a transcription activation domain. Preferably, the transcription regulating domain has at least 90% sequence identity to a mammalian, including human if the fusion protein is intended for human gene therapy, transcription regulating domain to avoid inducing undesirable immunological responses.

The transcription regulating domain can be any such domain known to regulator or prepared to regulate eukaryotic transcription. Such TRDs are known, and include, but are not limited to, VP16, VP64, TA2, STAT-6, p65, and derivatives, multimers and combinations thereof that exhibit transcriptional regulation properties. The transcription regulating domain can be derived from an intracellular receptor, such as a nuclear hormone receptor transcription activation (or repression) domain, and is preferably a steroid hormone receptor transcription activation domain or variant thereof that exhibits transcriptional regulation properties. Transcription domains include, but are not limited to, TAF-1, TAF-2, TAU-1, TAU-2, and variants thereof.

The transcription regulating domain may be a viral transcription activation domain or variant thereof. Preferably, the viral transcription regulating domain comprises a VP16 transcription activation domain or variant thereof.

The transcription regulating domain can include a transcription repression domain. Such domains are known, and include, but are not limited to, transcription repression domains selected from among ERD, KRAB, SID, Deacetylase, and derivatives, multimers and combinations thereof, such as KRAB-ERD, SID-ERD, (KRAB)₂, (KRAB)₃, KRAB-A, (KRAB-A)₂, (SID)₂ (KRAB-A)-SID and SID-(KRAB-A).

Nucleic acid constructs

Also provided are nucleic acid molecules that encode the resulting fusion proteins. The nucleic acids can be included in vectors, suitable for expression of the proteins and/or vectors suitable for gene therapy. Cell containing the vectors are also provided. Typically the cell is a eukaryotic cell. In other embodiments, the cell is a prokaryotic cell.

Also provided are expression cassettes that contain a gene of interest, particularly a gene encoding a therapeutic product, such as an angiogenesis inhibitor, operatively linked to a transcriptional regulatory region or response element, including sequences of nucleic acids to which a fusion proteins provided herein binds and controls transcription, particularly upon binding of a ligand to the LBD of the fusion polypeptide. Such expression cassettes can be included in a vector for gene therapy, and are intended for administration with, before or after, administration of the fusion protein or nucleic acid encoding the fusion protein. Genes of interest for exogenous delivery typically encode therapeutic proteins, such as growth factors, growth factor inhibitors or antagonists, tumor necrosis factor (TNF) inhibitors, anti-tumor agents, angiogenesis agents, anti-angiogenesis agents, clotting factors, apoptotic and other suicide genes.

Compositions, combinations and kits

Also provided are compositions that contain the fusion proteins or the vectors that encoded the fusion proteins. Combinations of the fusion proteins or nucleic acids encoding the proteins and nucleic acid encoding a targeted gene with regulatory regions selected for activation by the fusion protein are also provided.

Compositions, particularly pharmaceutical compositions containing the fusion polypeptides in a pharmaceutically acceptable carrier are also provided.

Combinations of the expression cassette and fusion polypeptide or nucleic acid molecules, particularly expression vectors that encode the fusion polypeptide are provided. The combinations may include separate

compositions or a single composition containing both elements. Kits containing the combinations and optionally instructions for administration thereof and other reagents used in preparing and administering the combinations are also provided.

5 Hence compositions suitable for gene therapy that contain nucleic acid encoding the fusion protein, typically in a vector suitable for gene therapy are provided. Preferred vectors include viral vectors, preferably adenoviral vectors, and lentiviral vectors. In other embodiments, non-viral delivery systems, including DNA-ligand complexes, adenovirus-ligand-DNA complexes, direct injection of DNA, CaPO₄ precipitation, gene gun techniques, electroporation, liposomes and lipofection are provided.

10

15 The compositions suitable for regulating gene expression contain an effective amount of the fusion protein or a polynucleotide encoding the ligand activated transcriptional regulatory fusion protein and a pharmaceutically acceptable excipient. Such compositions can further include a regulatable expression cassette encoding a gene and at least one response element for the gene recognized by the nucleotide binding domain of the fusion polypeptide.

20 The regulatable expression cassette is designed to include a sequence of nucleic acids with which the nucleic acid binding domain of the ligand activated transcriptional regulatory fusion protein interacts. It also preferably includes operatively linked transcriptional regulatory sequences that are regulatable by the TRD of the fusion protein. Typically, the regulatable expression cassette includes 3 to 6 response

25 elements.

Methods

30 Methods for regulating expression of endogenous and exogenous genes are provided. The methods are practiced by administering to a cell a composition that contains an effective amount or concentration of the fusion protein or of nucleic acid molecule, such as a vector that encodes the fusion protein. The nucleic acid binding domain (DBD) of the fusion

protein is selected to bind to a targeted nucleic acid sequence in the genome of the cell or in an exogenously administered nucleic acid molecule, and the transcription regulating domain (TRD) is selected to regulate transcription from a selected promoter, which typically is operatively linked to the targeted nucleic acid binding domain. The exogenously administered nucleic acid molecule comprises an expression cassette encoding a gene of interest and operatively linked to a regulatory region that contains elements, such as a promoter and response elements.

10 As noted the targeted regulatory region and gene of interest may be endogenously present in the cell or separately administered as part of an expression cassette encoding the gene of interest. If separately administered, it is administered as part of a regulatable expression cassette that includes a gene and at least one response element for the gene recognized by the nucleotide binding domain of the fusion protein.

15 At the same time or at a later time, a composition containing comprising a ligand that binds to the ligand binding domain of the fusion protein is also administered. The ligand can be administered in the same composition as the fusion protein (or encoding nucleic acid molecule) or in a separate composition. The ligand and fusion protein may be administered sequentially, simultaneously or intermittently.

20 Hence gene therapy is effected by administering a ligand that binds to the LBD of the fusion protein. Preferably the ligand is a non-natural ligand and the LBD has been modified from the native form present in native intracellular receptors to preferentially and selectively interact with the non-natural ligand. Upon administration, the ligand binds to the ligand binding domain of the fusion protein, whereby the DBD of the fusion protein, either as a monomer or dimer, interacts with a targeted gene and transcription of the targeted gene is repressed or activated. As noted, 25 the targeted gene may be an endogenous gene or an exogenously administered gene.

In other embodiments, the methods for regulating gene expression in a cell are effected by administering to the cell a composition containing an effective amount of the nucleic acid molecule that encodes the ligand activated transcriptional regulatory fusion protein, a regulatable expression cassette containing a gene operatively linked to at least one response element for the gene recognized by the nucleotide binding domain of the polypeptide encoded by the polynucleotide, and a pharmaceutically acceptable excipient; and administering to the cell a ligand that binds to the ligand binding domain of the encoded polypeptide, where the nucleotide binding domain of the encoded polypeptide binds to the response element and activates or represses transcription of the gene.

Methods for treating a cellular proliferative disorder by the *ex vivo* introduction of a recombinant expression vector encoding the fusion protein are provided. Cellular proliferative disorder include disorders associated with transcription of a gene at reduced or increased levels.

Administration can of the composition(s) can be effected *in vitro*, *in vivo* or *ex vivo*. One such method includes the removal of a tissue sample from a subject with a disorder, such as a cell proliferative disorder, isolating hematopoietic or other cells from the tissue sample, and contacting isolated cells with the fusion protein or a nucleic acid molecule encoding the fusion protein, and, optionally, a target specific gene. Optionally, the cells can be treated with a growth factor, such as interleukin-2 for example, to stimulate cell growth, before reintroducing the cells into the subject. When reintroduced, the cells specifically target the cell population from which they were originally isolated. In this way, the trans-repressing activity of the zinc finger-nucleotide binding polypeptide may be used to inhibit or suppress undesirable cell proliferation in a subject. Preferably, the subject is a human.

Results exemplified herein demonstrate ligand activated transcription of a targeted gene and demonstrate the utility of the fusion protein

containing a zinc finger DNA binding domain, such as a mammalian C2H2 DNA binding domain, a ligand binding domain from an intracellular receptor, such as an estrogen receptor, and, optionally, a heterologous transcription regulating domain for the purpose of obtaining ligand-dependent control of expression of a transgene introduced into mammalian cells. Hence it is shown herein that heterologous zinc finger domains can be combined with an intracellular receptor to achieve ligand-dependent gene expression of a targeted gene.

DESCRIPTION OF THE DRAWINGS

- 10 In the drawings, which form a portion of the specification:
- FIGURE 1 is a schematic for the selection strategy for the *in vitro* evolution of the 3 finger protein Zif268, recognizing its natural 9 bp target site (top), into a 6 finger protein, recognizing a desired 18 bp target sequence (bottom).
- 15 FIGURE 2 is a schematic depiction of the functional domains (A-F) of the human estrogen receptor.
- FIGURE 3 is a schematic depiction of the cloning strategy for the construction of the recombinant molecular constructs.
- 20 FIGURE 4 is a schematic map of the expression vector for C7LBDAS based on the plasmid pCDNA3.1.
- FIGURE 5 is a schematic map of the expression vector for C7LBDBS based on the plasmid pCDNA3.1.
- 25 FIGURE 6 is a schematic map of the expression vector for C7LBDCS based on the plasmid pCDNA3.1.
- FIGURE 7 is a schematic map of the expression vector for C7LBDAL based on the plasmid pCDNA3.1.
- 30 FIGURE 8 is a schematic map of the expression vector for C7LBDBL based on the plasmid pCDNA3.1.
- FIGURE 9 is a schematic map of the expression vector for C7LBDCL based on the plasmid pCDNA3.1.

FIGURE 10 is a schematic summary of the structure of several embodiments of the recombinant molecular construct and the nucleotide sequences of the DNA binding regions of zinc finger domains C7, E2C and 2C7.

5 FIGURE 11 is a schematic map of the expression vector for E2CLBDAS based on the plasmid pCDNA3.1.

FIGURE 12 is a schematic map of the expression vector for E2CLBDBS based on the plasmid pCDNA3.1.,

10 FIGURE 13 is a schematic diagram of the constructs C7LBDASTA2, C7LBDBSTA2, C7LBDBS-STAT6, C7LBDBSVP16 (SEQ ID NO: 16), AND C7LBDBSNLSVP16.

FIGURE 14 is a schematic restriction map of constructs comprising RXR and ecdysone (EcR) ligand binding domains used in heterodimers.

15 FIGURE 15 is a schematic depiction of the cloning strategy for the construction of the 2C7LBD recombinant molecular constructs.

FIGURE 16 is a schematic map of the expression vector for 2C7LBDAS based on the plasmid pCDNA3.1.

FIGURE 17 is a schematic map of the expression vector for 2C7LBDBS based on the plasmid pCDNA3.1.

20 FIGURE 18 is a schematic map of the expression vector for 2C7LBDCS based on the plasmid pCDNA3.1.

FIGURE 19 is a schematic map of the expression vector for LBDASNLSVP16 (SEQ ID NO: 13), based on the plasmid pCDNA3.1.

25 FIGURE 20 is a schematic map of the expression vector for C7LBDBSVP16 based on the plasmid pCDNA3.1.

FIGURE 21 is a schematic map of the expression vector for C7LBDBSG521R (SEQ ID NO: 15), based on the plasmid pCDNA3.1.

FIGURE 22 is a schematic map of the expression vector for C7LBDBSG400V (SEQ ID NO: 14), based on the plasmid pCDNA3.1.

30 FIGURE 23 shows A: an inducible promoter based on binding sites for the 3 Finger protein N1. The promoter contains 5 direct repeats of N1 sites

spaced by 3 bp; the spacing between the 5 repeats is 6 bp. Bottom: Luciferase assay. HeLa cells were cotransfected with plasmids encoding the indicated fusion proteins and the N1 reporter construct. Twenty four hours later, the cells were treated with 10 nM RU486 (B) or 100nM Tamoxifen, C respectively. Forty-eight hours post transfection, cell extracts were assayed for luciferase activity.

FIGURE 24 shows an inducible promoter based on binding sites for the 3 Finger protein B3. A: The promoter contains 5 direct repeats of B3 sites spaced by 3 bp; the spacing between the 5 repeats is 6 bp. Bottom: Luciferase assay. HeLa cells were cotransfected with plasmids encoding the indicated fusion proteins and the B3 reporter construct. At 24 h later, the cells were treated with 10 nM RU486 (B), or 100 nM Tamoxifen (C), respectively. At 48 h post transfection, cell extracts were assayed for luciferase activity.

FIGURE 25 is a graphical depiction of the results of luciferase assay showing the RU486-induced formation of functional VP64-C7-PR/VP64-CF2-PR heterodimers. HeLa cells were cotransfected with the corresponding effector plasmids and TATA reporter plasmids (C7/CF2-dr0, C7 site 5' to a CF2 site, direct "repeat", no spacing; C7/C7-dr0, 2 C7 sites, direct repeat, no spacing). At 24 h later, the cells were treated with 10 nM RU486. At 48 h post transfection, cell extracts were assayed for luciferase activity.

FIGURE 26 shows a restriction map for the plasmid designated pAvCVlx.

FIGURE 27 shows a restriction map for the plasmid designated pSQ3.

DETAILED DESCRIPTION

I. DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which this invention belongs. All patents, applications, published applications and other publications and sequences from

GenBank and other data bases referred to anywhere in the disclosure herein are incorporated by reference in their entirety.

As used herein, the ligand binding domain (LBD) of the fusion proteins provided herein refers to the portion of the fusion protein responsible for binding to a selected ligand. The LBD optionally and preferably includes dimerization and inactivation functions. The LBDs in the proteins herein are derived from the 300 amino acid carboxyl-terminal half of intracellular receptors, particularly those that are members of the steroid hormone nuclear receptor superfamily. It is the portion of the receptor protein with which a ligand interacts thereby inducing a cascade of events leading to the specific association of an activated receptor with regulatory elements of target genes. In these receptors the LDB includes the hormone binding function, the inactivation funciton, such as through interactions with heat shock proteins (hsp), and dimerization function.

The LBDs used herein include such LBDs and modified derivatives thereof, particularly forms with altered ligand specificity.

As used herein, the transcription regulating domain (TRD) refers to the portion of the fusion polypeptide provided herein that functions to regulate gene transcription. Exemplary and preferred transcription repressor domains are ERD, KRAB, SID, Deacetylase, and derivatives, multimers and combinations thereof such as KRAB-ERD, SID-ERD, (KRAB)₂, (KRAB)₃, KRAB-A, (KRAB-A)₂, (SID)₂ (KRAB-A)-SID and SID-(KRAB-A).

As used herein, the DNA binding domain (DBD), or alternatively the nucleic acid (or nucleotide) binding domain, refers to the portion of the fusion polypeptide provided herein that provides specific nucleic acid binding capability. The use of the abbreviation DBD is not meant to limit it to DNA binding domains, but is also intented to include polypeptides that bind to RNA. The nucleic acid binding domain functions to target the protein to specific genes by virtue of the specificity of the interaction of the TRD region for nucleotide sequences operatively linked to the

transcriptional apparatus of a gene. The DBD targets the fusion protein to the selected targeted gene or genes, which gene(s) may be endogenous or exogenously added.

As used herein, operatively linked means that elements of the fusion polypeptide, for example, are linked such that each perform or functions as intended. For example, the repressor is attached to the binding domain in such a manner that, when bound to a target nucleotide via that binding domain, the repressor acts to inhibit or prevent transcription. Linkage between and among elements may be direct or indirect, such as via a linker. The elements are not necessarily adjacent. Hence a repressor domain of a TRD can be linked to a DNA binding domain using any linking procedure well known in the art. It may be necessary to include a linker moiety between the two domains. Such a linker moiety is typically a short sequence of amino acid residues that provides spacing between the domains. So long as the linker does not interfere with any of the functions of the binding or repressor domains, any sequence can be used.

As used herein, a fusion protein is a protein that contains portions or fragments of two or more naturally-occurring proteins operatively joined or linked to form the fusion protein in which each fragment retains a function or a modified function exhibited by the naturally occurring proteins. The fragments from the naturally occurring protein may be modified to alter the original properties.

As used herein, modified, modification, mutant or other such terms refers to an alteration of the domain in question from its naturally occurring wild-type form, and includes primary sequence changes.

As used herein, "modulating" envisions the inhibition or suppression of expression from a promoter containing a zinc finger-nucleotide binding motif when it is over-activated, or augmentation or enhancement of expression from such a promoter when it is under-activated.

As used herein, steroid hormone receptor superfamily refers to the superfamily of intracellular receptors that are steroid receptors. Representative examples of such receptors include, but are not limited to, the estrogen, progesterone, glucocorticoid- α , glucocorticoid- β , mineralocorticoid, androgen, thyroid hormone, retinoic acid, retinoid X, Vitamin D, COUP-TF, ecdysone, Nurr-I and orphan receptors.

5 As used herein, the amino acids, which occur in the various amino acid sequences appearing herein, are identified according to their well-known, three-letter or one-letter abbreviations. The nucleotides, which occur in the various DNA fragments, are designated with the standard 10 single-letter designations used routinely in the art.

15 In a peptide or protein, suitable conservative substitutions of amino acids are known to those of skill in this art and may be made generally without altering the biological activity of the resulting molecule. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter 20 biological activity (see, e.g., Watson *et al.* *Molecular Biology of the Gene*, 4th Edition, 1987, The Bejacmin/Cummings Pub. co., p.224).

25 As used herein, a delivery plasmid is a plasmid vector that carries or delivers nucleotide acids encoding a therapeutic gene or gene that encodes a therapeutic product or a precursor thereof or a regulatory gene or other factor that results in a therapeutic effect when delivered *in vivo* in or into a cell line, such as, but not limited to a packaging cell line, to propagate therapeutic viral vectors.

30 As used herein, "recombinant expression vector" or "expression vector" refers to a plasmid, virus or other vehicle known in the art that has been manipulated by insertion or incorporation of heterologous DNA, such as nucleic acid encoding the fusion proteins herein or expression cassettes provided herein. Such expression vectors contain a promotor sequence for efficient transcription of the inserted nucleic acid in a cell. The expression vector typically contains an origin of replication, a

promoter, as well as specific genes that permit phenotypic selection of transformed cells.

As used herein, a DNA or nucleic acid homolog refers to a nucleic acid that includes a preselected conserved nucleotide sequence, such as a sequence encoding a therapeutic polypeptide. By the term "substantially homologous" is meant having at least 80%, preferably at least 90%, most preferably at least 95% homology therewith or a less percentage of homology or identity and conserved biological activity or function.

As used herein, "host cells" are cells in which a vector can be propagated and its DNA expressed. The term also includes any progeny of the subject host cell. It is understood that all progeny may not be identical to the parental cell since there may be mutations that occur during replication. Such progeny are included when the term "host cell" is used. Methods of stable transfer where the foreign DNA is continuously maintained in the host are known in the art.

The terms "homology" and "identity" are often used interchangeably. In this regard, percent homology or identity may be determined, for example, by comparing sequence information using a GAP computer program. The GAP program uses the alignment method of Needleman and Wunsch ((1970) *J. Mol. Biol.* 48:443), as revised by Smith and Waterman ((1981) *Adv. Appl. Math.* 2:482). Briefly, the GAP program defines similarity as the number of aligned symbols (i.e., nucleotides or amino acids) which are similar, divided by the total number of symbols in the shorter of the two sequences. The preferred default parameters for the GAP program may include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) and the weighted comparison matrix of Gribskov *et al.* (1986) *Nucl. Acids Res.* 14:6745, as described by Schwartz and Dayhoff, eds., *ATLAS OF PROTEIN SEQUENCE AND STRUCTURE*, National Biomedical Research Foundation, pp. 353-358 (1979); (2) a penalty of 3.0 for each gap and an additional 0.10 penalty for each symbol in each gap; and (3) no penalty for end gaps.

Whether any two nucleic acid molecules have nucleotide sequences that are at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% "identical" can be determined using known computer algorithms such as the "FAST A" program, using for example, the default parameters as in Pearson *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:2444. Alternatively the BLAST function of the National Center for Biotechnology Information database may be used to determine identity

In general, sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be calculated using published techniques. (See, e.g.: *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo *et al.* (1988) *SIAM J Applied Math* 48:1073). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in *Guide to Huge Computers*, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo *et al.* (1988) *SIAM J Applied Math* 48:1073. Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., *et al.*, *Nucleic Acids Research* 12(1):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al.*, *J Molec Biol* 215:403 (1990)).

Therefore, as used herein, the term "identity" represents a comparison between a test and a reference polypeptide or polynucleotide. For example, a test polypeptide may be defined as any polypeptide that is 90% or more identical to a reference polypeptide. As used herein, the term at least "90% identical to" refers to percent identities from 90 to 99.99 relative to the reference polypeptides. Identity at a level of 90% or more is indicative of the fact that, assuming for exemplification purposes a test and reference polynucleotide length of 100 amino acids are compared. No more than 10% (i.e., 10 out of 100) amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons may be made between a test and reference polynucleotides. Such differences may be represented as point mutations randomly distributed over the entire length of an amino acid sequence or they may be clustered in one or more locations of varying length up to the maximum allowable, e.g. 10/100 amino acid difference (approximately 90% identity). Differences are defined as nucleic acid or amino acid substitutions, or deletions.

As used herein, primer refers to an oligonucleotide containing two or more deoxyribonucleotides or ribonucleotides, preferably more than three, from which synthesis of a primer extension product can be initiated. For purposes herein, a primer of interest is one that is substantially complementary to a zinc finger-nucleotide binding protein strand, but also can introduce mutations into the amplification products at selected residue sites. Experimental conditions conducive to synthesis include the presence of nucleoside triphosphates and an agent for polymerization and extension, such as DNA polymerase, and a suitable buffer, temperature and pH.

As used herein, genetic therapy involves the transfer of heterologous DNA to the certain cells, target cells, of a mammal, particularly a human, with a disorder or conditions for which such therapy is sought. The DNA is introduced into the selected target cells in a

manner such that the heterologous DNA is expressed and a therapeutic product encoded thereby is produced. Alternatively, the heterologous DNA may in some manner mediate expression of DNA that encodes the therapeutic product, or it may encode a product, such as a peptide or RNA that in some manner mediates, directly or indirectly, expression of a therapeutic product. Genetic therapy may also be used to deliver nucleic acid encoding a gene product that replaces a defective gene or supplements a gene product produced by the mammal or the cell in which it is introduced. The introduced nucleic acid may encode a therapeutic compound, such as a growth factor inhibitor thereof, or a tumor necrosis factor or inhibitor thereor, such as a receptor therefor, that is not normally produced in the mammalian host or that is not produced in therapeutically effective amounts or at a therapeutically useful time. The heterologous DNA encoding the therapeutic product may be modified prior to introduction into the cells of the afflicted host in order to enhance or otherwise alter the product or expression thereof. Genetic therapy may also involve delivery of an inhibitor or repressor or other modulator of gene expression.

As used herein, heterologous DNA is DNA that encodes RNA and proteins that are not normally produced *in vivo* by the cell in which it is expressed or that mediates or encodes mediators that alter expression of endogenous DNA by affecting transcription, translation, or other regulatable biochemical processes. Heterologous DNA may also be referred to as foreign DNA. Any DNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which it is expressed is herein encompassed by heterologous DNA. Examples of heterologous DNA include, but are not limited to, DNA that encodes traceable marker proteins, such as a protein that confers drug resistance, DNA that encodes therapeutically effective substances, such as anti-cancer agents, enzymes and hormones, and DNA that encodes other types of proteins, such as antibodies. Antibodies that are encoded by

heterologous DNA may be secreted or expressed on the surface of the cell in which the heterologous DNA has been introduced.

Hence, herein heterologous DNA or foreign DNA, includes a DNA molecule not present in the exact orientation and position as the counterpart DNA molecule found in the genome. It may also refer to a DNA molecule from another organism or species (*i.e.*, exogenous).

As used herein, a therapeutically effective product is a product that is encoded by heterologous nucleic acid, typically DNA, that, upon introduction of the nucleic acid into a host, a product is expressed that ameliorates or eliminates the symptoms, manifestations of an inherited or acquired disease or that cures the disease.

Typically, DNA encoding a desired gene product is cloned into a plasmid vector and introduced by routine methods, such as calcium-phosphate mediated DNA uptake (see, (1981) *Somat. Cell. Mol. Genet.* 7:603-616) or microinjection, into producer cells, such as packaging cells.

After amplification in producer cells, the vectors that contain the heterologous DNA are introduced into selected target cells.

As used herein, an expression or delivery vector refers to any plasmid or virus into which a foreign or heterologous DNA may be inserted for expression in a suitable host cell — *i.e.*, the protein or polypeptide encoded by the DNA is synthesized in the host cell's system. Vectors capable of directing the expression of DNA segments (genes) encoding one or more proteins are referred to herein as "expression vectors." Also included are vectors that allow cloning of cDNA (complementary DNA) from mRNAs produced using reverse transcriptase.

As used herein, a gene refers to a nucleic acid molecule whose nucleotide sequence encodes an RNA or polypeptide. A gene can be either RNA or DNA. Genes may include regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).

As used herein, isolated with reference to a nucleic acid molecule or polypeptide or other biomolecule means that the nucleic acid or polypeptide has separated from the genetic environment from which the polypeptide or nucleic acid were obtained. It may also mean altered from the natural state. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Thus, a polypeptide or polynucleotide produced and/or contained within a recombinant host cell is considered isolated. Also intended as an "isolated polypeptide" or an "isolated polynucleotide" are polypeptides or polynucleotides that have been purified, partially or substantially, from a recombinant host cell or from a native source. For example, a recombinantly produced version of a compound can be substantially purified by the one-step method described in Smith *et al.* (1988) *Gene* 67:31-40. The terms isolated and purified are sometimes used interchangeably.

Thus, by "isolated" the nucleic acid is free of the coding sequences of those genes that, in a naturally-occurring genome immediately flank the gene encoding the nucleic acid of interest. Isolated DNA may be single-stranded or double-stranded, and may be genomic DNA, cDNA, recombinant hybrid DNA, or synthetic DNA. It may be identical to a native DNA sequence, or may differ from such sequence by the deletion, addition, or substitution of one or more nucleotides.

Isolated or purified as it refers to preparations made from biological cells or hosts means any cell extract containing the indicated DNA or protein including a crude extract of the DNA or protein of interest. For example, in the case of a protein, a purified preparation can be obtained following an individual technique or a series of preparative or biochemical techniques and the DNA or protein of interest can be present at various degrees of purity in these preparations. The procedures may include for example, but are not limited to, ammonium sulfate fractionation, gel

filtration, ion exchange change chromatography, affinity chromatography, density gradient centrifugation and electrophoresis.

A preparation of DNA or protein that is "substantially pure" or "isolated" should be understood to mean a preparation free from naturally occurring materials with which such DNA or protein is normally associated in nature. "Essentially pure" should be understood to mean a "highly" purified preparation that contains at least 95% of the DNA or protein of interest.

A cell extract that contains the DNA or protein of interest should be understood to mean a homogenate preparation or cell-free preparation obtained from cells that express the protein or contain the DNA of interest. The term "cell extract" is intended to include culture media, especially spent culture media from which the cells have been removed.

As used herein, "modulate" refers to the suppression, enhancement or induction of a function. For example, zinc finger-nucleic acid binding domains and variants thereof may modulate a promoter sequence by binding to a motif within the promoter, thereby enhancing or suppressing transcription of a gene operatively linked to the promoter cellular nucleotide sequence. Alternatively, modulation may include inhibition of transcription of a gene where the zinc finger-nucleotide binding polypeptide variant binds to the structural gene and blocks DNA dependent RNA polymerase from reading through the gene, thus inhibiting transcription of the gene. The structural gene may be a normal cellular gene or an oncogene, for example. Alternatively, modulation may include inhibition of translation of a transcript.

As used herein, "inhibit" refers to the suppression of the level of activation of transcription of a structural gene operably linked to a promoter. For example, for the methods herein the gene includes a zinc finger-nucleotide binding motif.

As used herein, a transcriptional regulatory region refers to a region that drives gene expression in the target cell. Transcriptional regulatory

regions suitable for use herein include but are not limited to the human cytomegalovirus (CMV) immediate-early enhancer/promoter, the SV40 early enhancer/promoter, the JC polyomavirus promoter, the albumin promoter, PGK and the α -actin promoter coupled to the CMV enhancer.

5 As used herein, a promoter region of a gene includes the regulatory elements that typically lie 5' to a structural gene. If a gene is to be activated, proteins known as transcription factors attach to the promoter region of the gene. This assembly resembles an "on switch" by enabling an enzyme to transcribe a second genetic segment from DNA
10 into RNA. In most cases the resulting RNA molecule serves as a template for synthesis of a specific protein; sometimes RNA itself is the final product. The promoter region may be a normal cellular promoter or, for example, an onco-promoter. An onco-promoter is generally a virus-derived promoter. Viral promoters to which zinc finger binding
15 polypeptides may be targeted include, but are not limited to, retroviral long terminal repeats (LTRs), and *Lentivirus* promoters, such as promoters from human T-cell lymphotropic virus (HTLV) 1 and 2 and human immunodeficiency virus (HIV) 1 or 2.

As used herein, "effective amount" includes that amount that results in the deactivation of a previously activated promoter or that amount that results in the inactivation of a promoter containing a zinc finger-nucleotide binding motif, or that amount that blocks transcription of a structural gene or translation of RNA. The amount of zinc finger derived-nucleotide binding polypeptide required is that amount necessary to either displace a native zinc finger-nucleotide binding protein in an existing protein/promoter complex, or that amount necessary to compete with the native zinc finger-nucleotide binding protein to form a complex with the promoter itself. Similarly, the amount required to block a structural gene or RNA is that amount which binds to and blocks RNA polymerase from reading through on the gene or that amount which inhibits translation, respectively. Preferably, the method is performed
20
25
30

intracellularly. By functionally inactivating a promoter or structural gene, transcription or translation is suppressed. Delivery of an effective amount of the inhibitory protein for binding to or "contacting" the cellular nucleotide sequence containing the zinc finger-nucleotide binding protein motif, can be accomplished by one of the mechanisms described herein, such as by retroviral vectors or liposomes, or other methods well known in the art.

As used herein, "truncated" refers to a zinc finger-nucleotide binding polypeptide derivative that contains less than the full number of zinc fingers found in the native zinc finger binding protein or that has been deleted of non-desired sequences. For example, truncation of the zinc finger-nucleotide binding protein TFIIIA, which naturally contains nine zinc fingers, might be a polypeptide with only zinc fingers one through three. Expansion refers to a zinc finger polypeptide to which additional zinc finger modules have been added. For example, TFIIIA may be extended to 12 fingers by adding 3 zinc finger domains. In addition, a truncated zinc finger-nucleotide binding polypeptide may include zinc finger modules from more than one wild type polypeptide, thus resulting in a "hybrid" zinc finger-nucleotide binding polypeptide.

As used herein, "mutagenized" refers to a zinc finger derived-nucleotide binding polypeptide that has been obtained by performing any of the known methods for accomplishing random or site-directed mutagenesis of the DNA encoding the protein. For instance, in TFIIIA, mutagenesis can be performed to replace nonconserved residues in one or more of the repeats of the consensus sequence. Truncated zinc finger-nucleotide binding proteins can also be mutagenized.

As used herein, a polypeptide "variant" or "derivative" refers to a polypeptide that is a mutagenized form of a polypeptide or one produced through recombination but that still retains a desired activity, such as the ability to bind to a ligand or a nucleic acid molecule or to modulate transcription.

As used herein, a zinc finger-nucleotide binding polypeptide "variant" or "derivative" refers to a polypeptide that is a mutagenized form of a zinc finger protein or one produced through recombination. A variant may be a hybrid that contains zinc finger domain(s) from one protein linked to zinc finger domain(s) of a second protein, for example. The domains may be wild type or mutagenized. A "variant" or "derivative" includes a truncated form of a wild type zinc finger protein, which contains less than the original number of fingers in the wild type protein. Examples of zinc finger-nucleotide binding polypeptides from which a derivative or variant may be produced include TFIIIA and zif268. Similar terms are used to refer to "variant" or "derivative" nuclear hormone receptors and "variant" or "derivative" transcription effector domains.

As used herein a "zinc finger-nucleotide binding motif" refers to any two or three-dimensional feature of a nucleotide segment to which a zinc finger-nucleotide binding derivative polypeptide binds with specificity. Included within this definition are nucleotide sequences, generally of five nucleotides or less, as well as the three dimensional aspects of the DNA double helix, such as, but are not limited to, the major and minor grooves and the face of the helix. The motif is typically any sequence of suitable length to which the zinc finger polypeptide can bind. For example, a three finger polypeptide binds to a motif typically having about 9 to about 14 base pairs. Preferably, the recognition sequence is at least about 16 base pairs to ensure specificity within the genome. Therefore, zinc finger-nucleotide binding polypeptides of any specificity are provided. The zinc finger binding motif can be any sequence designed empirically or to which the zinc finger protein binds. The motif may be found in any DNA or RNA sequence, including regulatory sequences, exons, introns, or any non-coding sequence.

As used herein, the terms "pharmaceutically acceptable", "physiologically tolerable" and grammatical variations thereof, as they

refer to compositions, carriers, diluents and reagents, are used interchangeably and represent that the materials are capable of administration to or upon a human without the production of undesirable physiological effects such as nausea, dizziness, gastric upset and the like which would be to a degree that would prohibit administration of the composition.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting between different genetic environments another nucleic acid to which it has been operatively linked. Preferred vectors are those capable of autonomous replication and expression of structural gene products present in the DNA segments to which they are operatively linked. Vectors, therefore, preferably contain the replicons and selectable markers described earlier.

As used herein with regard to nucleic acid molecules, including DNA fragments, the phrase "operatively linked" means the sequences or segments have been covalently joined, preferably by conventional phosphodiester bonds, into one strand of DNA, whether in single or double stranded form such that operatively linked portions functions as intended. The choice of vector to which transcription unit or a cassette provided herein is operatively linked depends directly, as is well known in the art, on the functional properties desired, *e.g.*, vector replication and protein expression, and the host cell to be transformed, these being limitations inherent in the art of constructing recombinant DNA molecules.

As used herein, a sequence of nucleotides adapted for directional ligation, *i.e.*, a polylinker, is a region of the DNA expression vector that (1) operatively links for replication and transport the upstream and downstream translatable DNA sequences and (2) provides a site or means for directional ligation of a DNA sequence into the vector. Typically, a directional polylinker is a sequence of nucleotides that defines two or more restriction endonuclease recognition sequences, or restriction sites. Upon restriction cleavage, the two sites yield cohesive termini to which a

translatable DNA sequence can be ligated to the DNA expression vector. Preferably, the two restriction sites provide, upon restriction cleavage, cohesive termini that are non-complementary and thereby permit directional insertion of a translatable DNA sequence into the cassette. In one embodiment, the directional ligation means is provided by nucleotides present in the upstream translatable DNA sequence, downstream translatable DNA sequence, or both. In another embodiment, the sequence of nucleotides adapted for directional ligation comprises a sequence of nucleotides that defines multiple directional cloning means.

Where the sequence of nucleotides adapted for directional ligation defines numerous restriction sites, it is referred to as a multiple cloning site.

As used herein, a secretion signal is a leader peptide domain of a protein that targets the protein to the periplasmic membrane of gram negative bacteria. A preferred secretion signal is a pelB secretion signal. The predicted amino acid residue sequences of the secretion signal domain from two pelB gene product variants from *Erwinia carotova* are described in Lei, *et al.* (*Nature*, 331:543-546, 1988). The leader sequence of the pelB protein has previously been used as a secretion signal for fusion proteins (Better *et al.* (1988) *Science* 240:1041-1043; Sastry *et al.* (1989) *Proc. Natl. Acad. Sci. USA* 86:5728-5732; and Mullinax *et al.* (1990) *Proc. Natl. Acad. Sci. USA*, 87:8095-8099). Amino acid residue sequences for other secretion signal polypeptide domains from *E. coli* are known (see, e.g., Oliver, In Neidhard, F.C. (ed.), *Escherichia coli and Salmonella Typhimurium*, American Society for Microbiology, Washington, D.C., 1:56-69 (1987)).

As used herein, ligand refers to any compound interacts with the ligand binding domain of a receptor and modulate its activity; ligands typically activate receptors. Ligand can also include compounds that activate the receptor without binding. A natural ligand is a compound that normally interacts with the receptor.

As used herein, anti-hormones are compounds that are antagonists of the naturally-occurring receptor. The anti-hormone is opposite in activity to a hormone.

As used herein, non-natural ligands or non-native ligands refer to compounds that are normally not found in mammals, such as humans, that bind to or interact with the ligand binding domain of a receptor. Hence, the term "non-native ligands" refers to those ligands that are not naturally found in the specific organism (man or animal) in which gene therapy is contemplated. For example, certain insect hormones such as ecdysone are not found in humans. As such ecdysone is non-native hormone to an animal, such as a human.

As used herein, "cell-proliferative disorder" denotes malignant as well as non-malignant disorders in which cell populations morphologically appear to differ from the surrounding tissue. The cell-proliferative disorder may be a transcriptional disorder that results in an increase or a decrease in gene expression level. The cause of the disorder may be of cellular origin or viral origin. Gene therapy using a zinc finger-nucleotide binding polypeptide can be used to treat a virus-induced cell proliferative disorder in a human, for example, as well as in a plant. Treatment can be prophylactic in order to make a plant cell, for example, resistant to a virus, or therapeutic, in order to ameliorate an established infection in a cell, by preventing production of viral products.

As used herein, "cellular nucleotide sequence" refers to a nucleotide sequence that is present within a cell. It is not necessary that the sequence be a naturally occurring sequence of the cell. For example, a retroviral genome that is integrated within a host's cellular DNA, would be considered a "cellular nucleotide sequence". The cellular nucleotide sequence can be DNA or RNA and includes introns and exons, DNA and RNA. The cell and/or cellular nucleotide sequence can be prokaryotic or eukaryotic, including a yeast, virus, or plant nucleotide sequence.

As used herein, administration of a therapeutic composition can be effected by any means, and includes, but is not limited to, subcutaneous, intravenous, intramuscular, intrasternal, infusion techniques, intraperitoneally administration and parenteral administration.

5 **II. Fusion Protein**

A. **General**

The fusion protein is constructed to include a ligand binding domain and a nucleic acid binding domain; the nucleic acid binding domain is not derived from the same receptor as the ligand binding domain. Inclusion of these two domains permits sequence specific binding to target nucleic acid sequences present in endogenous or exogenous nucleic acid molecules. It also provides ligand-dependent control of such sequence-specific binding. The fusion protein can also include a transcription regulating domain that serves to enhance, suppress or activate expression of an endogenous or exogenous gene. Such transcriptional control is also ligand dependent.

The nucleic acid binding domain (the DBD) includes one or more zinc finger peptide modular units, and typically a plurality of such units joined to provide a peptide designed to bind to the regulatory region in a targeted gene. Zinc fingers provide a means to design DBDs of a desired specificity.

The fusion protein also includes a LBD that derived from an intracellular receptor, preferably a hormone receptor, more preferably a steroid receptor. The LBD can be modified to have altered ligand specificity so that endogenous or natural ligands do not interact with it, but non-natural ligands do. The fusion protein also can include a transcription regulating domain (TRD) that regulates transcription of the targeted gene(s). In some embodiments, the TRD can repress transcription of an endogenous gene; in others it can activate expression of an endogenous or exogenous gene.

Hence the fusion protein is made by operably linking a LBD domain from an intracellular receptor to a one or more zinc finger domains, selected to bind to a targeted gene. A transcription regulating domain can also be operably linked. This is accomplished by any method known to those of skill in the art. Generally the fusion protein is produced by expressing nucleic acid encoding the fusion protein.

5 **1. Ligand Binding Domain (LBD)**

The ligand binding domain is derived from an intracellular receptor, and is preferably derived from a nuclear hormone receptor. The LBD of 10 an intracellular receptor includes the approximately 300 amino acids from the carboxy terminal, which can be used with or without modification.

15 By mutation of a small number of residues ligand specificity can be altered. The ligand binding domain can be modified, such as by truncation or point mutation to alter its ligand specificity permitting gene regulation by non-natural or non-native ligands.

20 Exemplary hormone receptors are steroid receptors, which are well known in the art. Exemplary and preferred steroid receptors include estrogen and progesterone receptors and variants thereof. Of particular interest are ligand binding domains that exhibit altered ligand specificity so that the LBD does not respond to the natural hormone, but rather to a drug, such as RU486, or other inducer. Means to modify and test the 25 specificity of ligand binding domains and to identify ligands therefor are known (see, U.S. Patent No. 5,874,534; U.S. Patent No. 5,935,934; and International PCT application No. 98/18925, which is based on U.S. provisional application Serial No. 60/029,964; International PCT application No. 96/40911, which is based on U.S. application Serial No. 08/479,913).

30 The LBD can be modified by deletion of from about 1 up to about 150, typically 120, amino acids on the carboxyl terminal end of the receptor from which the LBD derives. Systematic deletion of amino acids and subsequent testing of the ligand specificity and of the resulting LBD

can be used to empirically identify mutations that lead to modified LBDs that have desired properties, such as preferential interaction with non-natural ligands. Exemplary mutations are described in the Examples herein, and also are known to those of skill in the art (see, e.g., U.S. Patent No. 5,874,534; U.S. Patent No. 5,935,934; U.S. Patent No. 5,364,791; and International PCT application No. 98/18925, which is based on U.S. provisional application Serial No. 60/029,964; International PCT application No. 96/40911, which is based on U.S. application Serial No. 08/479,913) and references cited therein. Hence a LBD or modified form thereof prepared by known methods is obtained and operably linked to a DBD; a TRD is also linked as needed.

2. Nucleic Acid Binding Domain (DBD)

Zinc fingers are modular nucleic acid binding peptides. The zinc fingers, or modules thereof, or variant thereof can be used to construct fusion proteins that specifically interact with targeted sequences. Zinc fingers are ubiquitous proteins, and many are well-characterized. For example, methods and rules for preparation and selection of zinc fingers based upon the C2H2 class of zinc fingers with unique specificity are known (see, e.g., International PCT application No. WO 98/54311 and International PCT application No. 95/19431; see, also U.S. Patent No. 5,789,538; Beerli *et al.* (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96:2758-2763; Beerli *et al.* (1995) *Proc. Natl. Acad. Sci. U.S.A.* 95:14628-14633; see, also U.S. application Serial No. 09/173,941, filed 16 October, 1998, published as International PCT application No. WO 00/23464). Exemplary targeting sequences are provided herein.

Furthermore, other zinc fingers can be similarly identified and the rules known for the C2H2 can be applied to modification of the specificity of such zinc fingers or alternative rules unique to each class can be deduced in a similar manner.

The advantage of using zinc fingers for targeting of the ligand-dependent transcription regulating fusion proteins provided herein is the

ability to construct zinc fingers with unique specificity. This permits targeting and ligand-dependent control of expression of specific endogenous genes and also ligand-dependent control of exogenously administered genes, such as genes that encode therapeutic products.

5 Zinc fingers and modular units thereof can be obtained or prepared by any method known to those of skill in the art. As discussed herein, a plethora of zinc fingers, including synthetic zinc fingers having a variety of sequence specificities are known, as are means for combining the modular domains to produce a resulting peptide that binds to any desired
10 target sequence of nucleic acids. Rules for creating zinc fingers of desired specificity are known and can be deduced by methods used by those of skill in the art (see, e.g., (see, e.g., International PCT application No. WO 98/54311, which is based on U.S. application Serial No. 08/863,813; International PCT application No. 95/19431, which is based
15 on U.S. application Serial Nos. 08/183,119 and 08/312,604).

For example, zinc finger variants can be prepared by identifying a zinc finger or modular unit thereof, creating an expression library, such as a phage display library (see, e.g., International PCT application No. WO 98/54311, Barbas *et al.* (1991) *Methods* 2:119; Barbas *et al.* (1992)
20 *Proc. Natl. Acad. Sci. U.S.A.* 89:4457), encoding polypeptide variants of the zinc finger or modular unit therof, expressing the library in a host and screening for variant peptides having a desired specificity. Zinc fingers may also be constructed by combining amino acids (or encoding nucleic acids) according to the known rules of binding specificity and, if
25 necessary, testing or screening the resulting peptides to ensure the peptide has a desired specificity. Because of the modular nature of zinc fingers, where each module can be prepared to bind to three nucleotide sequence, peptides of any specificity can be prepared from the modules. The number of modules used depends upon the specificity of gene
30 targeting desired. Modular units are combined; spacers (*i.e.* TGEKP,

TGQKP) required to maintain spacing and conformational features of the modular domains are included in the peptide (see, e.g., WO 98/54311).

5 a. **Zinc fingers as DBDs and zinc finger modular units**

The nucleic acid binding domain in the fusion protein includes zinc finger modular domains and is designed to bind to a target nucleic acid sequence present in an endogenous gene or in an exogenous gene that is administered in combination with the fusion protein or nucleic acid encoding the fusion protein.

10 Zinc fingers are among the most common and ubiquitous nucleic acid binding proteins. Any zinc finger polypeptide or modular unit thereof is contemplated; preferably the domain is non-immunogenic in the host for which the fusion protein is intended. For human therapy, the zinc finger DBD preferably is selected from human zinc protein modular units or variants thereof.

15 For purposes herein, the zinc finger used generally is other than the naturally-occurring zinc finger present in the intracellular receptor from which the ligand binding domain is derived. Typically the fusion protein is produced by replacing the native zinc finger present in the receptor with the selected zinc finger designed to interact with a targeted nucleic acid regulatory region. In addition, the zinc fingers can be designed by selection of appropriate modular units to have specificity for a targeted gene, thereby providing a precise means to modulate expression of a targeted gene.

20 Naturally occurring zinc finger proteins generally contain multiple repeats of the zinc finger motif. This modular nature is unique among the different classes of DNA binding proteins. Wild type zinc finger proteins are made up of from two to as many as 37 modular tandem repeats, with each repeat forming a "finger" holding a zinc atom in tetrahedral coordination by means of a pair of conserved cysteines and a pair of conserved histidines. Generally each finger also contains conserved hydrophobic amino acids that interact to form a hydrophobic core that

helps the module maintain its shape. Polydactyl arrays of as many as 37 zinc finger domains allow this recognition domain to recognize extended asymmetric sequences. Any such zinc finger or combinations of modular units thereof is intended for use herein.

5 A zinc finger-nucleotide binding peptide domain contains a unique heptamer (contiguous sequence of 7 amino acid residues) within the α -helical domain of the polypeptide, which heptameric sequence determines binding specificity to a target nucleotide. The heptameric sequence can be located anywhere within the α -helical domain but it is preferred that the heptamer extend from position -1 to position 6 as the residues are conventionally numbered in the art. A peptide nucleotide-binding domain can include any β -sheet and framework sequences known in the art to function as part of a zinc finger protein.

10 Studies of natural zinc finger proteins have shown that three zinc finger domains can bind 9 bp of contiguous DNA sequence (Pavletich *et al.* (1991) *Science* 252:809-817; Swirnoff *et al.* (1995) *Mol. Cell. Biol.* 15:2275-2287). While recognition of 9 bp of sequence is insufficient to specify a unique site in a complex genome, proteins containing six zinc finger domains can specify 18-bp recognition (Liu *et al.* (1997) *Proc. Natl. Acad. Sci. USA* 94:5525-5530). An 18-bp address made up of modular units is of sufficient complexity to specify a single site within all known genomes (see, published International PCT application No. WO 98/54311). Rules for constructing Zinc finger arrays that bind to a particular DNA sequence are known (see, e.g., International PCT application No. WO 98/54311, which is based on U.S. application Serial No. 08/863,813; International PCT application No. 95/19431, which is based on U.S. application Serial Nos. 08/183,119 and 08/312,604).

15 20 25 30 Zinc finger-nucleotide binding polypeptide variants can be constructed from known motifs. The variants include at least two and preferably at least about four zinc finger modules that bind to a cellular

nucleotide sequence, such as DNA, RNA or both, and specifically bind to and modulate the function of a cellular nucleotide sequence.

For purposes herein, it is not necessary that the zinc finger-nucleotide binding motif be known in order to obtain a zinc-finger nucleotide binding variant polypeptide. It is contemplated that zinc finger-nucleotide binding motifs can be identified in non-eukaryotic DNA or RNA, especially in the native promoters of bacteria and viruses by the binding thereto of the modified nucleic acid binding peptides. Modified nucleic acid binding peptides should preserve the well known structural characteristics of the zinc finger, but differ from zinc finger proteins found in nature by their amino acid sequences and three-dimensional structures.

A variety of zinc finger proteins are known. Among these, the Cys₂-His₂ (also referred to as "C2H2") zinc fingers are preferred for use in the fusion proteins. There are well-defined rules for C2H2 zinc finger binding to DNA that allow the DNA binding specificity of the fusion proteins containing the zinc fingers to be adjusted in order to reduce non-specific interactions with genes other than the targeted genes. These proteins can be selected or engineered to bind to diverse sequences. Further, the sequence specificity of these proteins can be modified to be different from their naturally occurring targets. Examples of zinc finger proteins from which a polypeptide can be produced include TFIIIA and Zif268.

The murine Cys₂-His₂ zinc finger protein Zif268 has been used for construction of phage display libraries (Wu *et al.* (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92:344-348). Zif268 is structurally the most well characterized of the zinc-finger proteins (Pavletich, *et al.* (1991) *Science* 252:809-817; Elrod-Erickson *et al.* (1996) *Structure* 4:1171-1180; Swirnoff *et al.* (1995) *Mol. Cell. Biol.* 15:2275-2287). DNA recognition in each of the three zinc finger domains of this protein is mediated by residues in the N-terminus of the α -helix contacting primarily three nucleotides on a single strand of the DNA. The operator binding site for

this three finger protein is 5'-GCGTGGGCG-'3 (finger-2 subsite is underlined). Structural studies of Zif268 and other related zinc finger-DNA complexes have shown that residues from primarily three positions on the α -helix, -1, 3, and 6, are involved in specific base contacts.

5 Typically, the residue at position -1 of the α -helix contacts the 3' base of that finger's subsite while positions 3 and 6 contact the middle base and the 5' base, respectively.

b. Construction and isolation of zinc finger DBD peptides

10 A zinc finger-nucleotide binding polypeptide that binds to DNA, and specifically, the zinc finger domains that bind to DNA, can be identified by examination of the "linker" region between two zinc finger domains. The linker amino acid sequence TGEK(P) (SEQ ID NO: 19) is typically indicative of zinc finger domains that bind to a DNA. Therefore, one can determine whether a particular zinc finger-nucleotide binding polypeptide preferably binds to DNA or RNA by examination of the linker amino acids.

c. Synthetic zinc fingers

20 Synthetic zinc fingers can be assembled based upon known sequence specificities. A large number of zinc finger-nucleotide binding polypeptides were made and tested for binding specificity against target nucleotides containing a GNN triplet. The data show that a striking conservation of all three of the primary DNA contact positions (-1, 3, and 6) was observed for virtually all the clones of a given target (see, Example 1, see, also U.S. application Serial No. 09/173,941, filed 16 October, 1998, published as International PCT application No. WO 00/23464).

25 In order to select a family of zinc finger domains recognizing the 5'-GNN-3' subset of sequences, two highly diverse zinc finger libraries were constructed in the phage display vector pComb3H (Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:7978-7982; Rader *et al.* (1997) *Curr. Opin. Biotechnol.* 8:503-508). Both libraries involved randomization of residues within the α -helix of finger 2 of C7, a variant of Zif268 (Wu *et al.* (1995) *Proc. Natl. Acad. Sci. U.S.A.* 92:344-348). Library 1 was

constructed by randomization of positions -1,1,2,3,5,6 using a NNK doping strategy while library 2 was constructed using a VNS doping strategy with randomization of positions -2,-1,1,2,3,5,6. The NNK doping strategy allows for all amino acid combinations within 32 codons while VNS precludes Tyr, Phe, Cys and all stop codons in its 24 codon set. The libraries contained 4.4×10^9 and 3.5×10^9 members, respectively, each capable of recognizing sequences of the 5'-GCGNNNGCG-3' type. The size of the NNK library ensured that it could be surveyed with 99% confidence while the VNS library was highly diverse but somewhat incomplete. These libraries are, however, significantly larger than previously reported zinc finger libraries.

(International PCT application No. WO 09/54311; Choo *et al.* (1994) *Proc Natl Acad Sci U S A* 91:11163-7; Greisman *et al.* (1997) *Science* 275:657-661; Rebar *et al.* (1994) *Science* 263:671-673; Jamieson *et al.* (1994) *Biochemistry* 33:5689-5695; Jamieson *et al.* 1996) *Proc. Natl. Acad. Sci. U.S.A.* 93:12834-12839; Isalan *et al.* (1998) *Biochemistry* 37:12026-12033; and U.S. Patent No. 5,789,538). Seven rounds of selection were performed on the zinc finger displaying-phage with each of the 16 5'-GCGNNNGCG-3' biotinylated hairpin DNAs targets using a solution binding protocol. Stringency was increased in each round by the addition of competitor DNA. Sheared herring sperm DNA was provided for selection against phage that bound non-specifically to DNA. Stringent selective pressure for sequence specificity was obtained by providing DNAs of the 5'-GCGNNNGCG-3' types as specific competitors. Excess DNA of the 5'-GCGNNNGCG-3' type was added to provide even more stringent selection against binding to DNAs with single or double base changes as compared to the biotinylated target. Phage binding to the single biotinylated DNA target sequence were recovered using streptavidin coated beads. In some cases the selection process was repeated. The data show that these domains are functionally modular and can be recombined with one another to create proteins capable of

binding to 18-bp sequences with subnanomolar affinity. The resulting family of zinc finger domains described herein is sufficient for the construction of 17 million proteins that bind to the 5'-(GNN)₆-3' family of DNA sequences.

5 Also impressive amino acid conservation was been observed for recognition of the same nucleotide in different targets. For example, Asn in position 3 (Asn3) virtually always selects to recognize adenine in the middle position, whether in the context of GAG, GAA, GAT, or GAC. Gln-1 and Arg-1 were always selected to recognize adenine or guanine, respectively, in the 3' position regardless of context. Amide side chain based recognition of adenine by Gln or Asn is well documented in structural studies as is the Arg guanidinium side chain to guanine contact with a 3' or 5' guanine (see, e.g., Elrod-Erickson *et al.* (1998) *Structure* 6:451-464).

10 More often, however, two or three amino acids are selected for nucleotide recognition. His3 or Lys3 (and to a lesser extent, Gly3) are selected for the recognition of a middle guanine. Ser3 and Ala3 are selected to recognize a middle thymine. Thr3, Asp3, and Glu3 are selected to recognize a middle cytosine. Asp and Glu were are selected in position -1 to recognize a 3' cytosine, while Thr-1 and Ser-1 are selected to recognize a 3' thymine.

15 More often, however, two or three amino acids are selected for nucleotide recognition. His3 or Lys3 (and to a lesser extent, Gly3) are selected for the recognition of a middle guanine. Ser3 and Ala3 are selected to recognize a middle thymine. Thr3, Asp3, and Glu3 are selected to recognize a middle cytosine. Asp and Glu were are selected in position -1 to recognize a 3' cytosine, while Thr-1 and Ser-1 are selected to recognize a 3' thymine.

20 Specific recognition of many nucleotides can best accomplished using motifs, rather than a single amino acid. For example, the best specification of a 3' guanine is achieved using the combination of Arg-1, Ser1, and Asp2 (the RSD motif). By using Val5 and Arg6 to specify a 5' guanine, recognition of subsites GGG, GAG, GTG, and GCG can be accomplished using a common helix structure (SRSD-X-LVR) differing only in the position 3 residue (Lys3 for GGG, Asn3 for GAG, Glu3 for GTG, and Asp3 for GCG). Similarly, 3' thymine is specified using Thr-1, Ser1, and Gly2 in the final clones(the TSG motif). Further, a 3' cytosine can be specified using Asp-1, Pro1, and Gly2 (the DPG motif) except when the

subsite is GCC; Pro1 is not tolerated by this subsite. Specification of a 3' adenine is with Gln-1, Ser1, Ser2 in two clones (QSS motif).

The data (see, Table 1 in Example) show that all possible GNN triplet sequences can be recognized with exquisite specificity by zinc finger domains. Optimized zinc finger domains can discriminate single base differences by greater than 100-fold loss in affinity. While many of the amino acids found in the optimized proteins at the key contact positions -1,3, and 6 are those that are consistent with a simple code of recognition, it has been discovered that optimal specific recognition is sensitive to the context in which these residues are presented. Residues at positions 1,2, and 5 have been found to be critical for specific recognition.

Further the data demonstrate that sequence motifs at positions -1,1, and 2 rather than the simple identity of the position 1 residue are required for highly specific recognition of the 3' base. These residues likely provide the proper stereo-chemical context for interactions of the helix in terms of recognition of specific bases and in the exclusion of other bases, the net result being highly specific interactions. Ready recombination of the disclosed domains then allows for the creation of proteins, typically polypdactyl proteins, of defined specificity precluding the need to develop phage display libraries in their generation. Such family of zinc finger domains is sufficient for the construction of 16 or 17 million proteins that bind to the 5'-(GNN)₆-3' family of DNA sequences.

d. Modification of zinc finger peptides

The zinc finger-nucleotide binding peptide domain can be derived or produced from a wild type zinc finger protein by truncation or expansion, or as a variant of the wild type-derived polypeptide by a process of site directed mutagenesis, or by a combination of the procedures (see, e.g., U.S. Patent No. 5,789,538, which describes methods for design and construction of zinc finger peptides). Mutagenesis can be performed to replace non-conserved residues in one or more of the repeats of the

consensus sequence. Truncated zinc finger-nucleotide binding proteins can also be mutagenized.

DNA encoding the zinc finger-nucleotide binding proteins, including native, truncated, and expanded polypeptides, can be obtained by several methods. For example, the DNA can be isolated using hybridization procedures which are well known in the art. These include, but are not limited to: (1) hybridization of probes to genomic or cDNA libraries to detect shared nucleotide sequences; (2) antibody screening of expression libraries to detect shared structural features; and (3) synthesis by the polymerase chain reaction (PCR). RNA can be obtained by methods known in the art (seem e.g., *Current Protocols in Molecular Biology*, 1988, Ed. Ausubel, *et al.*, Greene Publish. Assoc. & Wiley Interscience).

DNA encoding zinc finger-nucleotide binding proteins also can be obtained by: (1) isolation of a double-stranded DNA sequence from the genomic DNA; (2) chemical manufacture of a DNA sequence to provide the necessary codons for the polypeptide of interest; and (3) *in vitro* synthesis of a double-stranded DNA sequence by reverse transcription of mRNA isolated from a eukaryotic donor cell. In the latter case, a double-stranded DNA complement of mRNA is eventually formed which is generally referred to as cDNA. Of these three methods the isolation of genomic DNA is the least common. This is especially true when it is desirable to obtain the microbial expression of mammalian polypeptides due to the presence of introns.

For obtaining zinc finger derived-DNA binding polypeptides, the synthesis of DNA sequences is frequently the method of choice when the entire sequence of amino acid residues of the desired polypeptide product is known. When the entire sequence of amino acid residues of the desired polypeptide is not known, the direct synthesis of DNA sequences is not possible and the method of choice is the formation of cDNA sequences. Among the standard procedures for isolating cDNA sequences of interest is the formation of plasmid-carrying cDNA libraries

which are derived from reverse transcription of mRNA which is abundant in donor cells that have a high level of genetic expression. When used in combination with polymerase chain reaction technology, even rare expression products can be cloned. In those cases where significant portions of the amino acid sequence of the polypeptide are known, the production of labeled single or double-stranded DNA or RNA probe sequences duplicating a sequence putatively present in the target cDNA may be employed in DNA/DNA hybridization procedures which are carried out on cloned copies of the cDNA which have been denatured into a single-stranded form (*Jay, et al., Nucleic Acid Research, 11:2325, 1983*).

Hybridization procedures are useful for the screening of recombinant clones by using labeled mixed synthetic oligonucleotide probes where each probe is potentially the complete complement of a specific DNA sequence in the hybridization sample which includes a heterogeneous mixture of denatured double-stranded DNA. For such screening, hybridization is preferably performed on either single-stranded DNA or denatured double-stranded DNA. Hybridization is particularly useful in the detection of cDNA clones derived from sources where an extremely low amount of mRNA sequences relating to the polypeptide of interest are present. By using stringent hybridization conditions directed to avoid non-specific binding, it is possible, for example, to allow the autoradiographic visualization of a specific cDNA clone by the hybridization of the target DNA to that single probe in the mixture which is its complete complement (*Wallace, et al., Nucleic Acid Research, 9:879, 1981; Maniatis, et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, 1982*).

Screening procedures that rely on nucleic acid hybridization make it possible to isolate any gene sequence from any organism, provided the appropriate probe is available. Oligonucleotide probes, which correspond to a part of the sequence encoding the protein in question, can be synthesized chemically. This requires that short, oligopeptide stretches of

amino acid sequence must be known. The DNA sequence encoding the protein can be deduced from the genetic code, however, the degeneracy of the code must be taken into account. It is possible to perform a mixed addition reaction when the sequence is degenerate. This includes a heterogeneous mixture of denatured double-stranded DNA. For such screening, hybridization is preferably performed on either single-stranded DNA or denatured double-stranded DNA.

A cDNA expression library, such as lambda gt11, can be screened indirectly for zinc finger-nucleotide binding protein or for the zinc finger derived polypeptide having at least one epitope, using antibodies specific for the zinc finger-nucleotide binding protein. Such antibodies can be either polyclonally or monoclonally derived and used to detect expression product indicative of the presence of zinc finger-nucleotide binding protein cDNA. Alternatively, binding of the derived polypeptides to DNA targets can be assayed by incorporated radiolabeled DNA into the target site and testing for retardation of electrophoretic mobility as compared with unbound target site.

A preferred vector used for identification of truncated and/or mutagenized zinc finger-nucleotide binding polypeptides is a recombinant DNA molecule containing a nucleotide sequence that codes for and is capable of expressing a fusion polypeptide containing, in the direction of amino- to carboxy-terminus, (1) a prokaryotic secretion signal domain, (2) a heterologous polypeptide, and (3) a filamentous phage membrane anchor domain. The vector includes DNA expression control sequences for expressing the fusion polypeptide, preferably prokaryotic control sequences.

Since the DNA sequences provided herein encode essentially all or part of an zinc finger-nucleotide binding protein, it is routine to prepare, subclone, and express the truncated polypeptide fragments of DNA from this or corresponding DNA sequences. Alternatively, by using the DNA fragments disclosed herein, which define the zinc finger-nucleotide

binding polypeptides, it is possible, in conjunction with known techniques, to determine the DNA sequences encoding the entire zinc finger-nucleotide binding protein. Such techniques are described in U.S. 4,394,443 and U.S. 4,446,235, which are incorporated herein by reference.

In addition to modifications in the amino acids making up the zinc finger, the zinc finger derived polypeptide can contain more or less than the full amount of fingers contained in the wild type protein from which it is derived. Minor modifications of the primary amino acid sequence may result in proteins which have substantially equivalent activity compared to the zinc finger derived-binding protein described herein. Such modifications may be deliberate, as by site-directed mutagenesis, or may be spontaneous. All proteins produced by these modifications are included herein as long as zinc finger-nucleotide binding protein activity exists.

e. Screening of variant zinc finger and other DBD peptides

Any method known to those of skill in the art for identification of functional modular domains derived from zinc fingers and combinations thereof can be employed. An exemplary method for identifying variants of zinc fingers or other polypeptides that bind to zinc finger binding motifs is provided. Components used in the method include a nucleic acid molecule encoding a putative or modified zinc finger peptide operably linked to a first inducible promoter and a reporter gene operably linked to a second inducible promoter and a zinc finger-nucleotide binding motif, wherein the incubating is carried out under conditions sufficient to allow the components to interact, and measuring the effect of the putative DBD peptide on the expression of the reporter gene is provided.

For example, a first inducible promoter, such as the arabinose promoter, is operably linked to the nucleotide sequence encoding the putative DBD polypeptide. A second inducible promoter, such as the

lactose promoter, is operably linked to a zinc finger derived-DNA binding motif followed by a reporter gene, such as β -galactosidase. Incubation of the components may be *in vitro* or *in vivo*. *In vivo* incubation may include prokaryotic or eukaryotic systems, such as *E.coli* or COS cells,

5 respectively. Conditions that allow the assay to proceed include incubation in the presence of a substance, such as arabinose and lactose, which activate the first and second inducible promoters, respectively, thereby allowing expression of the nucleotide sequence encoding the putative trans-modulating protein nucleotide sequence. Determination of

10 whether the putative modulating protein binds to the zinc finger-nucleotide binding motif, which is operably linked to the second inducible promoter, and affects its activity is measured by the expression of the reporter gene. For example, if the reporter gene is β -galactosidase, the presence of blue or white plaques indicates whether the putative modulating protein enhances or inhibits, respectively, gene expression from the promoter. Other commonly used assays to assess the function from a promoter, including chloramphenicol acetyl transferase (CAT) assay, are known to those of skill in the art. Prokaryote and eukaryote systems can be used.

15

20 As discussed above, Example 1 provides an illustration of modification of Zif268 as described above. Therefore, in another embodiment, a ligand activated transcriptional regulator polypeptide variant containing at least two zinc finger modules that bind to an HIV sequence and modulates the function of the HIV sequence, for example, the HIV promoter sequence is provided.

25

30 In another embodiment, zinc finger proteins can be manipulated to recognize and bind to extended target sequences. For example, zinc finger proteins containing from about 2 to 20 zinc fingers Zif(2) to Zif(20), and preferably from about 2 to 12 zinc fingers, may be fused to the leucine zipper domains of the Jun/Fos proteins, prototypical members of the bZIP family of proteins (O'Shea *et al.* (1991) *Science* 254:539).

Alternatively, zinc finger proteins can be fused to other proteins which are capable of forming heterodimers and contain dimerization domains. Such proteins are known to those of skill in the art.

The Jun/Fos leucine zippers are described for illustrative purposes and preferentially form heterodimers and allow for the recognition of 12 to 72 base pairs. Henceforth, Jun/Fos refer to the leucine zipper domains of these proteins. Zinc finger proteins are fused to Jun, and independently to Fos by methods commonly used in the art to link proteins. Following purification, the Zif-Jun and Zif-Fos constructs, the proteins are mixed to spontaneously form a Zif-Jun/Zif-Fos heterodimer. Alternatively, coexpression of the genes encoding these proteins results in the formation of Zif-Jun/Zif-Fos heterodimers *in vivo*. Fusion of the heterodimer with an N-terminal nuclear localization signal allows for targeting of expression to the nucleus (Calderon, *et al.*, *Cell*, **41**:499; 1982). Activation domains may also be incorporated into one or each of the leucine zipper fusion constructs to produce activators of transcription (Sadowski *et al.* (1992) *Gene* **118**:137). These dimeric constructs then allow for specific activation or repression of transcription. These heterodimeric Zif constructs are advantageous since they allow for recognition of palindromic sequences (if the fingers on Jun and Fos recognize the same DNA/RNA sequence) or extended asymmetric sequences (if the fingers on Jun and Fos recognize different DNA/RNA sequences). For example the palindromic sequence

5' - GGC CCA CGC {N}_x GCG TGG GCG - 3'

3' - GCG GGT GCG {N}_x CGC ACC CGC - 5' (SEQ ID NO: 20)

is recognized by the Zif268-Fos/Zif268 Jun dimer (x is any number). The spacing between subsites is determined by the site of fusion of Zif with the Jun or Fos zipper domains and the length of the linker between the Zif and zipper domains. Subsite spacing is determined by a binding site selection method as is common to those skilled in the art (Thiesen *et al.* (1990) *Nucleic Acids Research*, **18**:3203, 1990). Example of the

recognition of an extended asymmetric sequence is shown by the Zif(C7)₆-Jun/Zif-268-Fos dimer. This protein includes 6 fingers of the C7 type (EXAMPLE 11) linked to Jun and three fingers of Zif268 linked to Fos, and recognizes the extended sequence:

5 5' - CGC CGC CGC CGC CGC CGC {N}_x GCG TGG GCG - 3'
 3' - GCG GCG GCG GCG GCG GCG {N}_x CGC ACC CGC - 5'
 (SEQ ID NO: 21)

In another embodiment, attachment of chelating groups to Zif proteins is preferably facilitated by the incorporation of a Cysteine (Cys) residue between the initial Methionine (Met) and the first Tyrosine (Tyr) of the protein. The Cys is then alkylated with chelators known to those skilled in the art, for example, EDTA derivatives as described (Sigman (1990) *Biochemistry*, **29**:9097). Alternatively the sequence Gly-Gly-His can be made as the most amino terminal residues since an amino terminus composed of the residues has been described to chelate Cu⁺² (Mack *et al.* (1988) *J. Am. Chem. Soc.* **110**:7572). Preferred metal ions include Cu⁺², Ce⁺³ (Takasaki and Chin (1994) *J. Am. Chem. Soc.* **116**:1121, 1994) Zn⁺², Cd⁺², Pb⁺², Fe⁺² (Schnaith *et al.* (1994) *Proc. Natl. Acad. Sci., USA* **91**:569, 1994), Fe⁺³, Ni⁺², Ni⁺³, La⁺³, Eu⁺³ (Hallet *et al.* (1994) *Chemistry and Biology* **1**:185), Gd⁺³, Tb⁺³, Lu⁺³, Mn⁺², Mg⁺². Cleavage with chelated metals is generally performed in the presence of oxidizing agents such as O₂, hydrogen peroxide H₂O₂ and reducing agents such as thiols and ascorbate. The site and strand (+ or - site) of cleavage is determined empirically (Mack *et al.* (1988) *J. Am. Chem. Soc.* **110**:7572, 1988) and is dependent on the position of the Cys between the Met and the Tyr preceding the first finger. In the protein Met (AA) Tyr-(Zif)₁₋₁₂, the chelate becomes Met-(AA)_{x1} Cys-Chelate-(AA)_{x2}-Tyr-(Zif)₁₋₁₂, where AA = any amino acid and x = the number of amino acids. Dimeric zif constructs of the type Zif-Jun/Zif-Fos are preferred for cleavage at two sites within the target oligonucleotide or at a single long target site. In the case where double stranded cleavage is desired, Jun

and Fos containing proteins are labelled with chelators and cleavage is performed by methods known to those skilled in the art. In this case, a staggered double-stranded cut analogous to that produced by restriction enzymes is generated.

5 Following mutagenesis and selection of variants of the Zif268 protein in which the finger 1 specificity or affinity is modified, proteins carrying multiple copies of the finger may be constructed using the TGEKP linker sequence by methods known in the art. For example, the C7 finger may be constructed according to the scheme:

10 MKLLEPYACPVESCDRRFSKSADLKRHIRHTGEKP- (SEQ ID NO: 22)
(YACPVESCDRRFSKSADLKHIRHTGEKP) ₁₋₁₁, (SEQ ID NO: 23) where the sequence of the last linker is subject to change since it is at the terminus and not involved in linking two fingers together. This protein binds the designed target sequence GCG-GCG-GCG in the oligonucleotide hairpin CCT-CGC-CGC-CGC-GGG-TTT-TCC-CGC-GCC-CCC GAG G (SEQ ID NO: 24) with an affinity of 9nM, as compared to an affinity of 300 nM for an oligonucleotide encoding the GCG-TGG-GCG sequence (as determined by surface plasmon resonance studies). Fingers used need not be identical and may be mixed and matched to produce proteins which recognize a desired target sequence. These may also be used with leucine zippers (e.g., Fos/Jun) or other heterodimers to produce proteins with extended sequence recognition.

20 In addition to producing polymers of finger 1, the entire three finger Zif268 and modified versions therein may be fused using the consensus linker TGEKP to produce proteins with extended recognition sites. For example, the protein Zif268-Zif268 can be produced in which the natural protein has been fused to itself using the TGEKP linker. This protein now binds the sequence GCG-TGG-GCG-GCG-TGG-GCG. Therefore modifications within the three fingers of Zif268 or other zinc finger 25 proteins known in the art may be fused together to form a protein which

recognizes extended sequences. These new zinc proteins may also be used in combination with leucine zippers if desired.

3. Transcription regulating domain (TRD)

Any TRD known to those of skill in the art can be selected,
5 including those present in intracellular receptors. The TRD is selected to regulate transcription of the gene targeted by the DBD and to effect regulation of expression thereof. The TRD can be selected to regulate expression of an endogenous gene in a cell or in an exogenously added construct. For exogenously added genes, the regulatory region of the
10 gene can be selected to interact with a desired TRD. Identification, preparation and testing of TRDs in combination with DBDs is exemplified herein for ERB-2 and integrin β_3 .

a. Selection of the TRD

Transcription regulating domains are well known in the art.
15 Exemplary and preferred transcription repressor domains are ERD, KRAB, SID, Deacetylase, and derivatives, multimers and combinations thereof such as KRAB-ERD, SID-ERD, (KRAB)₂, (KRAB)₃, KRAB-A, (KRAB-A)₂, (SID)₂ (KRAB-A)-SID and SID-(KRAB-A).

b. Repressors

Transcriptional repressors are well known in the art, and any such repressor can be used herein. The repressor is a polypeptide that is operatively linked to the nucleic acid binding domain as set forth above.
20 The repressor is operatively linked to the binding domain in that it is attached to the binding domain in such a manner that, when bound to a target nucleotide via that binding domain, the repressor acts to inhibit or prevent transcription. The repressor domain can be linked to the binding domain using any linking procedure well known in the art. It may be necessary to include a linker moiety between the two domains. Such a linker moiety is typically a short sequence of amino acid residues that
25 provides spacing between the domains. So long as the linker does not
30

interfere with any of the functions of the binding or repressor domains, any sequence can be used.

Transcriptional repressors have been generated by attaching either of three human-derived repressor domains to the zinc finger protein. The first repressor protein was prepared using the ERF repressor domain (ERD) (Sgouras *et al.* (1995) *EMBO J.* 14:4781-4793), defined by amino acids 473 to 530 of the ets2 repressor factor (ERF). This domain mediates the antagonistic effect of ERF on the activity of transcription factors of the *ets* family. A synthetic repressor was constructed by fusion of this domain to the C-terminus of the zinc finger protein.

The second repressor protein was prepared using the Krüppel-associated box (KRAB) domain (Margolin *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:4509-4513). This repressor domain is commonly found at the N-terminus of zinc finger proteins and presumably exerts its repressive activity on TATA-dependent transcription in a distance- and orientation-independent manner (Pengue *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:1015-1020), by interacting with the RING finger protein KAP-1 (Friedman *et al.* (1996) *Genes & Dev.* 10:2067-2078). The KRAB domain found between amino acids 1 and 97 of the zinc finger protein KOX1 (Margolin *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:4509-4513) was used. In this case an N-terminal fusion with the six-finger protein was constructed.

Histone deacetylation as a means for repression can be employed. For example, amino acids 1 to 36 of the Mad mSIN3 interaction domain (SID) have been fused to the N-terminus of a zinc finger protein (Ayer *et al.* (1996) *Mol. Cell. Biol.* 16:5772-5781). This small domain is found at the N-terminus of the transcription factor Mad and is responsible for mediating its transcriptional repression by interacting with mSIN3, which in turn interacts the co-repressor N-CoR and with the histone deacetylase mRPD1 (Heinzel *et al.* (1997) *Nature* 387:43-46).

5 c. Activators

Exemplary and preferred transcription activation domains include any protein or factor that regulates transcription. Exemplary transcriptional regulation domains include, but are not limited to, VP16, TA2, VP64, STAT6 and relA.

10 4. **Exemplary construct based on human integrin β 3 and erbB-2 target sequences**

To exemplify the generation of zinc finger modular domains and peptides containing one or more of such domains to produce peptides with DNA binding specificity and therapeutic potential, target sequences have been identified based on human integrin β 3 and erbB-2 (Ishii *et al.* (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84:4374-4378) genomic sequences.

15 *Integrin β 3 as a target for cancer gene therapy*

Integrin $\alpha_v\beta_3$ is the most promiscuous member of the integrin family and has been identified as a marker of angiogenic vascular tissue. For instance, integrin $\alpha_v\beta_3$ shows enhanced expression on blood vessels in human wound granulation tissue but not in normal skin. Following the induction of angiogenesis, blood vessels show a four-fold increase in $\alpha_v\beta_3$ expression compared to blood vessels not undergoing this process. It has been reported that a cyclic peptide or monoclonal antibody antagonist of integrin $\alpha_v\beta_3$ blocks cytokine- or tumor-induced angiogenesis on the chick chorioallantoic membrane. Therefore, inhibition of integrin $\alpha_v\beta_3$ expression provides an approach to block tumor-induced angiogenesis.

20 *ErbB-2 receptor tyrosine kinases as a target for cancer gene therapy*

Members of the ErbB receptor family play an important role in the development of human malignancies. In particular, ErbB-2 is over-expressed as a result of gene amplification and/or transcriptional deregulation in a high percentage of human adenocarcinomas arising at numerous sites, including breast, ovary, lung, stomach, and salivary gland. Increased expression of ErbB-2 *per se* leads to constitutive activation of its intrinsic tyrosine kinase. Many clinical studies have

shown that patients with tumors showing elevated expression of ErbB-2 have poorer prognosis. Thus, the high occurrence of its aberrant expression in human cancer, as well as the aggressive behavior of over-expressing tumors, make ErbB-2 an attractive target for therapy.

5 Generation and construction of zinc fingers and fusion proteins targeted to *erbB-2* and integrin β_3 are described in the EXAMPLES.

B. Regulatable cassette

In embodiments in which the targeted gene is an exogenous gene, particularly a gene that encodes a therapeutic product, the gene is provided as in an expression cassette operatively linked to a promoter and regulatory region with which the fusion protein specifically interacts. The cassette includes at least one polynucleotide domain recognized by the corresponding zinc finger domain present in the fusion protein and a suitable promoter to direct transcription of the exogenous gene.

15 Typically, the regulatable expression cassette contains three to six response elements and interacts with nucleic acid binding domain of the ligand activated transcriptional regulatory fusion protein.

Typically the exogenous gene encodes a therapeutic product, such as a growth factor, that can supplement peptides, polypeptides or proteins encoded by endogenous expressed genes, thereby providing an effective therapy. In several embodiments the gene encodes a suitable reporter molecule that can be detected by suitable direct or indirect means. The cassette can be inserted into a suitable delivery vehicle for introduction into cells. Such vehicles include, but are not limited to, 20 human adenovirus vectors, adeno-associated vectors, murine or lenti virus derived retroviral vectors, and a variety of non-viral compositions including liposomes, polymers, and other DNA containing conjugates.

C. Use of the fusion proteins for gene regulation

1. Delivery of the nucleic acids

30 There are available to one skilled in the art multiple viral and non-viral methods suitable for introduction of a nucleic acid molecule into a

target cell. Genetic modification of a cell may be accomplished using one or more techniques well known in the gene therapy field (Human Gene Therapy, April 1994, Vol. 5, p. 543-563; Mulligan, R.C. 1993).

5 The ability to regulate transgene expression, as defined in the examples herein, can be applied to a wide variety of applications for gene therapy. The ability to control expression of an exogenously introduced transgene is important for the safety and efficacy of most or all envisioned cell and gene therapies. Control of transgene expression can be used to accomplish regulation of a therapeutic protein level, ablation of 10 a desired-cell population, either the vector containing cells or others, or activation of a recombinase or other function resulting in control of vector function within the transduced cells. Further, such control permits termination of a gene therapy treatment if necessary.

15 A number of vector systems useful for gene therapy have been described previously in this application. Vectors for gene therapy include any known to those of skill in the art, and include any vectors derived from animal viruses and artificial chromosomes. The vectors may be designed for integration into the host cell's chromosomes or to remain as extrachromosomal elements. Such vectors include, but are not limited to 20 human adenovirus vectors, adeno-associated viral vectors, retroviral vectors, such as murine retroviral vectors and lentivirus-derived retroviral vectors. Also contemplated herein are any of the variety of non-viral compositions for targeting and/or delivery of genetic material, including, but are not limited to, liposomes, polymers, and other DNA containing 25 compositions, and targeted conjugates, such as nucleic acids linked to antibodies and growth factors. Any delivery system is intended for use of delivery of the nucleic acid constructs encoding the fusion polypeptide and also targeted exogenous genes. Such vector systems can be used to deliver the ZFP-LBD fusion proteins and the inducible transgene cassette 30 either in vitro or in vivo, depending on the vector system. With adenovirus, for instance, vectors can be administered intravenously to

transduce the liver and other organs, introduced directly into the lung, or into vascular compartments temporarily localized by ligation or other methods. Methods for constructing such vectors, and methods and uses thereof are known to those skilled in the field of gene therapy.

5 In one embodiment, one vector encodes the fusion protein regulator and a second vector encodes the inducible transgene cassette. Vectors can be mixed or delivered sequentially to incorporate into cells the regulator and transgene at the appropriate amounts. Subsequent administration of an effective amount of the ligand by standard routes would result in activation of the transgene.

10 In another embodiment, the nucleic acid encoding the fusion protein and the inducible transgene can be included in the same vector construction. In this instance, the nucleic acid encoding the fusion protein would be positioned within the vector and expressed from a promoter in such a way that it did not interfere with the basal expression and inducibility of the transgene cassette. Further, the use of cell or tissue specific promoters to express the fusion protein confers an additional level of specificity on the system. Dual component vectors and use for gene therapy are known (see, e.g., Burcin *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96: 335-360, which describes an adenovirus vector fully deleted of viral backbone genes).

15 In another embodiment, gene therapy can be accomplished using a combination of the vectors described above. For example, a retroviral vector can deliver a stably integrated, inducible transgene cassette into a population of cells either *in vitro* (*ex vivo*) or *in vivo*. Subsequently, the integrated transgene can be activated by transducing this same cell population with a second vector, such as an adenovirus vector capable of expressing the fusion protein, followed by the administration of the specific ligand inducing agent. This is particularly useful where "one time" activation of the transgene is desired, for example as a cellular suicide mechanism. An example of this application is the stable

integration of an inducible transgene cassette containing the herpes simplex virus thymidine kinase gene (HSV Tk). Subsequent activation of this gene confers sensitivity to ganciclovir and allows ablation of this modified cell.

5 a. **Viral Delivery systems**

Viral transduction methods for delivering nucleic acid constructs to cells are contemplated herein. Suitable DNA viral vectors for use herein includes, but are not limited to an adenovirus (Ad), adeno-associated virus (AAV), herpes virus, vaccinia virus or a polio virus. A suitable RNA virus for use herein includes but is not limited to a retrovirus or Sindbis virus. It is to be understood by those skilled in the art that several such DNA and RNA viruses exist that may be suitable for use herein. Adenoviral vectors have proven especially useful for gene transfer into eukaryotic cells and are widely available to one skilled in the art and is suitable for use herein.

15 Adeno-associated virus (AAV) has recently been introduced as a gene transfer system with potential applications in gene therapy. Wild-type AAV demonstrates high-level infectivity, broad host range and specificity in integrating into the host cell genome. Herpes simplex virus type-1 (HSV-1) vectors are available and are especially useful in the nervous system because of its neurotropic property. Vaccinia viruses, of the poxvirus family, have also been developed as expression vectors. Each of the above-described vectors is widely available and is suitable for use herein.

25 Retroviral vectors are capable of infecting a large percentage of the target cells and integrating into the cell genome. Preferred retroviruses include lentiviruses, such as but are not limited to, HIV, BIV and SIV.

30 Various viral vectors that can be used for gene therapy as taught herein include adenovirus, herpes virus, vaccinia, adeno-associated virus (AAV), or, preferably, an RNA virus such as a retrovirus. Preferably, the retroviral vector is a derivative of a murine or avian retrovirus, or is a

lentiviral vector. The preferred retroviral vector is a lentiviral vector. Examples of retroviral vectors in which a single foreign gene can be inserted include, but are not limited to: Moloney murine leukemia virus (MoMuLV), Harvey murine sarcoma virus (HaMuSV), murine mammary tumor virus (MuMTV), SIV, BIV, HIV and Rous Sarcoma Virus (RSV). A number of additional retroviral vectors can incorporate multiple genes. All of these vectors can transfer or incorporate a gene for a selectable marker so that transduced cells can be identified and generated. By inserting a zinc finger derived-DNA binding polypeptide sequence of interest into the viral vector, along with another gene that encodes the ligand for a receptor on a specific target cell, for example, the vector is made target specific. Retroviral vectors can be made target specific by inserting, for example, a polynucleotide encoding a protein. Preferred targeting is accomplished by using an antibody to target the retroviral vector. Those of skill in the art know of, or can readily ascertain without undue experimentation, specific polynucleotide sequences which can be inserted into the retroviral genome to allow target specific delivery of the retroviral vector containing the zinc finger-nucleotide binding protein polynucleotide.

Since recombinant retroviruses are defective, they require assistance in order to produce infectious vector particles. This assistance can be provided, for example, by using helper cell lines that contain plasmids encoding all of the structural genes of the retrovirus under the control of regulatory sequences within the LTR. These plasmids are missing a nucleotide sequence which enables the packaging mechanism to recognize an RNA transcript for encapsitation. Helper cell lines which have deletions of the packaging signal include but are not limited to Ψ 2, PA317 and PA12, for example. These cell lines produce empty virions, since no genome is packaged. If a retroviral vector is introduced into such cells in which the packaging signal is intact, but the structural genes are replaced by other genes of interest, the vector can be packaged and

vector virion produced. The vector virions produced by this method can then be used to infect a tissue cell line, such as NIH 3T3 cells, to produce large quantities of chimeric retroviral virions.

5 **b. Nonviral Delivery systems**

"Non-viral" delivery techniques for gene therapy include DNA-ligand complexes, adenovirus-ligand-DNA complexes, direct injection of DNA, CaPO₄ precipitation, gene gun techniques, electroporation, liposomes and lipofection. Any of these methods are available to one skilled in the art and would be suitable for use herein. Other suitable methods are 10 available to one skilled in the art, and it is to be understood that the herein may be accomplished using any of the available methods of transfection.

Another targeted delivery system is a colloidal dispersion system. Colloidal dispersion systems include macromolecule complexes, 15 nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes, which are preferred. Liposomes are artificial membrane vesicles which are useful as delivery vehicles *in vitro* and *in vivo*. It has been shown that large unilamellar vesicles (LUV), which range in size from 0.2-4.0 μm can 20 encapsulate a substantial percentage of an aqueous buffer containing large macromolecules. RNA, DNA and intact virions can be encapsulated within the aqueous interior and be delivered to cells in a biologically active form (Fraley, *et al.*, *Trends Biochem. Sci.*, 6:77, 1981).

Lipofection may be accomplished by encapsulating an isolated 25 nucleic acid molecule within a liposomal particle and contacting the liposomal particle with the cell membrane of the target cell. Liposomes are self-assembling, colloidal particles in which a lipid bilayer, composed of amphiphilic molecules such as phosphatidyl serine or phosphatidyl choline, encapsulates a portion of the surrounding media such that the 30 lipid bilayer surrounds a hydrophilic interior. Unilamellar or multilamellar liposomes can be constructed such that the interior

contains a desired chemical, drug, or, as provide herein, an isolated nucleic acid molecule.

Liposomes have been used for delivery of polynucleotides in plant, yeast and bacterial cells as well as mammalian cells. In order for a liposome to be an efficient gene transfer vehicle, characteristics among the following should be present: (1) encapsulation of the genes of interest at high efficiency while not compromising their biological activity; (2) preferential and substantial binding to a target cell in comparison to non-target cells; (3) delivery of the aqueous contents of the vesicle to the target cell cytoplasm at high efficiency; and (4) accurate and effective expression of genetic information (Mannino, *et al.*, *Biotechniques*, 6:682, 1988).

The composition of the liposome is usually a combination of phospholipids, particularly high-phase-transition-temperature phospholipids, usually in combination with steroids, especially cholesterol. Other phospholipids or other lipids may also be used. The physical characteristics of liposomes depend on pH, ionic strength, and the presence of divalent cations.

Examples of lipids useful in liposome production include phosphatidyl compounds, such as phosphatidylglycerol, phosphatidylcholine, phosphatidylserine, phosphatidylethanolamine, sphingolipids, cerebrosides, and gangliosides. Particularly useful are diacylphosphatidylglycerols, where the lipid moiety contains from 14-18 carbon atoms, particularly from 16-18 carbon atoms, and is saturated. Illustrative phospholipids include egg phosphatidylcholine, dipalmitoylphosphatidylcholine and distearoylphosphatidylcholine.

The targeting of liposomes has been classified based on anatomical and mechanistic factors. Anatomical classification is based on the level of selectivity, for example, organ-specific, cell-specific, and organelle-specific. Mechanistic targeting can be distinguished based upon whether it is passive or active. Passive targeting uses the natural tendency of

liposomes to distribute to cells of the reticulo-endothelial system (RES) in organs which contain sinusoidal capillaries. Active targeting, on the other hand, involves alteration of the liposome by coupling the liposome to a specific ligand such as a monoclonal antibody, sugar, glycolipid, or protein, or by changing the composition or size of the liposome in order to achieve targeting to organs and cell types other than the naturally occurring sites of localization.

The surface of the targeted delivery system may be modified in a variety of ways. In the case of a liposomal targeted delivery system, lipid groups can be incorporated into the lipid bilayer of the liposome in order to maintain the targeting ligand in stable association with the liposomal bilayer. Various linking groups can be used for joining the lipid chains to the targeting ligand.

In general, the compounds bound to the surface of the targeted delivery system are ligands and receptors permitting the targeted delivery system to find and "home in" on the desired cells. A ligand may be any compound of interest that interacts with another compound, such as a receptor.

In general, surface membrane proteins that bind to specific effector molecules are referred to as receptors. Antibodies are preferred receptors. Antibodies can be used to target liposomes to specific cell-surface ligands. For example, certain antigens expressed specifically on tumor cells, referred to as tumor-associated antigens (TAAs), may be exploited for the purpose of targeting antibody-zinc finger-nucleotide binding protein-containing liposomes directly to the malignant tumor. Since the zinc finger-nucleotide binding protein gene product may be indiscriminate with respect to cell type in its action, a targeted delivery system offers a significant improvement over randomly injecting non-specific liposomes. A number of procedures can be used to covalently attach either polyclonal or monoclonal antibodies to a liposome bilayer. Antibody-targeted liposomes can include monoclonal or polyclonal

antibodies or fragments thereof such as Fab, or F(ab')₂, as long as they bind efficiently to an the antigenic epitope on the target cells. Liposomes may also be targeted to cells expressing receptors for hormones or other serum factors.

5 **2. Administration**

a. **Delivery of constructs to cells**

The cells may be transfected *in vivo*, *ex vivo* or *in vitro*. The cells may be transfected as primary cells isolated from a patient or a cell line derived from primary cells, and are not necessarily autologous to the patient to whom the cells are ultimately administered. Following *ex vivo* or *in vitro* transfection, the cells may be implanted into a host. Genetic modification of the cells may be accomplished using one or more techniques well known in the gene therapy field (see, e.g., (1994) *Human Gene Therapy* 5:543-563).

15 Administration of a nucleic acid molecules provided herein to a target cell *in vivo* may be accomplished using any of a variety of techniques well known to those skilled in the art. The vectors of the herein may be administered orally, parentally, by inhalation spray, rectally, or topically in dosage unit formulations containing conventional pharmaceutical acceptable carriers, adjuvants, and vehicles.

20 Suppositories for rectal administration of the drug can be prepared by mixing the drug with a suitable non-irritating excipient such as cocoa butter and polyethylene glycols that are solid at ordinary temperatures but liquid at the rectal temperature and therefore melt in the rectum and release the drug.

25 The dosage regimen for treating a disorder or a disease with the vectors and/or compositions provided is based on a variety of factors, including the type of disease, the age, weight, sex, medical condition of the patient, the severity of the condition, the route of administration, and the particular compound employed. Thus, the dosage regimen may vary widely, but can be determined empirically using standard methods.

The pharmaceutically active compounds (i.e., vectors) can be processed in accordance with conventional methods of pharmacy to produce medicinal agents for administration to patients, including humans and other mammals. For oral administration, the pharmaceutical composition may be in the form of, for example, a capsule, a tablet, a suspension, or liquid. The pharmaceutical composition is preferably made in the form of a dosage unit containing a given amount of DNA or viral vector particles (collectively referred to as "vector"). For example, these may contain an amount of vector from about 10^3 - 10^{15} viral vector particles, preferably from about 10^6 - 10^{12} viral particles. A suitable daily dose for a human or other mammal may vary widely depending on the condition of the patient and other factors, but, once again, can be determined using routine methods. The vector may also be administered by injection as a composition with suitable carriers including saline, dextrose, or water.

While the nucleic acids and /or vectors herein can be administered as the sole active pharmaceutical agent, they can also be used in combination with one or more vectors or other agents. When administered as a combination, the therapeutic agents can be formulated as separate compositions that are given at the same time or different times, or the therapeutic agents can be given as a single composition.

b. Deliver ligand

Ligands similarly may be delivered by any suitable mode of administration, including by oral, parenteral, intravenous, intramuscular and other known routes. Any known pharmaceutical formulations is contemplated.

3. Ligands

As noted, the ligands may be naturally-occurring ligands, but are preferentially non-natural ligands with which the LBD is modified to specifically interact. Methods for modifying the LBD are known, as are methods for screening for such ligands.

Ligands include, non-natural ligands, hormones, anti-hormones, synthetic hormones, and other such compounds. Examples of non-natural ligands, anti-hormones and non-native ligands include, but are not limited to, the following: 11 β -4-dimethylaminophenyl)-17 α -hydroxy-17 α -propinyl-4,9-e stradiene-3-one (RU38486 or Mifepestone); 11 β -(4-dimethylaminophenyl)-17 α -hydroxy-17 β -(3-hydroxypropyl)-13 α -methyl-4,9-gonadiene-3-one (ZK98299 or Onapristone); 11 β -(4-acetylphenyl)-17 β -hydroxy-17 α -(1-propinyl)-4,9-estradiene-3-one (ZK112993); 11 β -(4-dimethylaminophenyl)-17 β -hydroxy-17 α -(3-hydroxy-1 (Z)-propenyl-estra-4,9-diene-3-one (ZK98734); (7 β 11 β ,17 β)-11-(4-dimethylaminophenyl)-7-methyl-4',5'-dihydrospiroy'ester-4,9-diene-17,2' (3'H)-furan-3-one (Org31806); (11 β ,14 β ,17 α)-4',5'-dihydro-11-(4-dimethylamino-phenyl)spiroestra-4,9-diene-17,2'(3'H)-furan-3-one (Org31376); 5-alpha-pregnane-3,2-dione. Additional non-natural ligands include, in general, synthetic non-steroidal estrogenic or anti-estrogenic compounds, broadly defined as selective estrogen receptor modulators (SERMS). Exemplary compounds include, but are not limited to, tamoxifen and raloxifene.

4. Pharmaceutical compositions and combinations

Also provided is a pharmaceutical composition containing a therapeutically effective amount of the fusion protein, or a nucleic acid molecule encoding the fusion protein in a pharmaceutically acceptable carrier. Pharmaceutical compositions containing one or more fusion proteins with different zinc finger-nucleotide binding domains are contemplated. Also provided are pharmaceutical compositions containing the expression cassettes, and also compositions containing the ligands. Combinations containing a plurality of compositions are also provided.

Preparation of the compositions

The preparation of a pharmacological composition that contains active ingredients dissolved or dispersed therein is well known. Typically such compositions are prepared as sterile injectables either as liquid

solutions or suspensions, aqueous or non-aqueous, however, solid forms suitable for solution, or suspensions, in liquid prior to use can also be prepared. The preparation can also be emulsified. Tablets and other solid forms are contemplated.

5 The active ingredient can be mixed with excipients that are pharmaceutically acceptable and compatible with the active ingredient and in amounts suitable for use in the therapeutic methods described herein. Suitable excipients are, for example, water, saline, dextrose, glycerol, ethanol or the like and combinations thereof. In addition, if
10 desired, the composition can contain minor amounts of auxiliary substances such as wetting or emulsifying agents, as well as pH buffering agents and the like which enhance the effectiveness of the active ingredient.

15 The therapeutic pharmaceutical composition can include pharmaceutically acceptable salts of the components therein. Pharmaceutically acceptable salts include the acid addition salts (formed with the free amino groups of the polypeptide) that are formed with inorganic acids such as, for example, hydrochloric or phosphoric acids, or such organic acids as acetic, tartaric, mandelic and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, 2-ethylamino ethanol, histidine, procaine and others.

20 Physiologically tolerable carriers are well known in the art.
25 Exemplary of liquid carriers are sterile aqueous solutions that contain no materials in addition to the active ingredients and water, or contain a buffer such as sodium phosphate at physiological pH value, physiological saline or both, such as phosphate-buffered saline. Still further, aqueous carriers can contain more than one buffer salt, as well as salts such as sodium and potassium chlorides, dextrose, propylene glycol, polyethylene glycol and other solutes.

Liquid compositions can also contain liquid phases in addition to and to the exclusion of water. Exemplary of such additional liquid phases are glycerin, vegetable oils such as cottonseed oil, organic esters such as ethyl oleate, and water-oil emulsions.

5 D. Methods of gene regulation

Method of regulating expression of endogenous and exogenous genes are provided. In particular, ligand-dependent methods are provided. In practicing the methods, a target nucleotide acid molecule containing a sequence that interacts with the nucleic acid binding domain of the fusion protein exposed to an effective amount of the fusion protein in the presence of an effective binding amount of a ligand, which can be added simultaneous with or subsequent to the fusion protein. The nucleic acid binding domain of the fusion protein binds to a portion of the target nucleic acid molecule and the ligand binds to the ligand binding domain of the fusion protein. Exposure can occur *in vitro*, *in situ* or *in vivo*.

The amount of zinc finger derived-nucleotide binding polypeptide required is that amount necessary to either displace a native zinc finger-nucleotide binding protein in an existing protein/promoter complex, or that amount necessary to compete with the native zinc finger-nucleotide binding protein to form a complex with the promoter itself. Similarly, the amount required to block a structural gene or RNA is that amount which binds to and blocks RNA polymerase from reading through on the gene or that amount which inhibits translation, respectively. Preferably, the method is performed intracellularly. By functionally inactivating a promoter or structural gene, transcription or translation is suppressed. Delivery of an effective amount of the inhibitory protein for binding to or "contacting" the cellular nucleotide sequence containing the zinc finger-nucleotide binding protein motif, can be accomplished by one of the mechanisms described herein, such as by retroviral vectors or liposomes, or other methods well known in the art.

In one embodiment, a method for inhibiting or suppressing the function of a cellular gene or regulatory sequence that includes a zinc finger-nucleotide binding motif. This is effected by contacting the zinc finger-nucleotide binding motif with an effective amount of a fusion protein that includes zinc finger-nucleotide binding polypeptide derivative that binds to the motif. In instances in which the cellular nucleotide sequence is a promoter, the method includes inhibiting the transcriptional transactivation of a promoter containing a zinc finger-DNA binding motif. The zinc finger-nucleotide binding polypeptide derivative may bind to a motif within a structural gene or within an RNA sequence.

Treatments

Methods for gene therapy are provided. The fusion proteins are administered either as a protein or as a nucleic acid encoding the protein and delivered to cells or tissues in a mammal, such as a human. The fusion protein is targeted either to a specific sequence in the genome (an endogenous gene) or to an exogenously added gene, which is administered as part of an expression cassette. Prior to, simultaneous with or subsequent to administration of the fusion protein, a ligand that specifically interacts with the LBD in the fusion protein is administered. In embodiments, in which the targeted gene is exogenous, the expression cassette, which can be present in a vector, is administered, simultaneous with or subsequent to administration of the fusion protein. These methods are intended for treatment of any genetic disease, for treatment of acquired disease and any other conditions. Diseases include, cell proliferative disorders, such as cancer. Such therapy achieves its therapeutic effect by introduction of the fusion protein that includes the zinc finger-nucleotide binding polypeptide, either as the fusion or protein or encoded by a nucleic acid molecule that is expressed in the cells, into cells of animals having the disorder. Delivery of the fusion protein or nucleic acid molecule can be effected by any method known to those of skill in the art, including methods described herein. For example, it can

be effected using a recombinant expression vector such as a chimeric virus or a colloidal dispersion system.

The fusion proteins provided herein can be used for treating a variety of disorders. For example the proteins can be used for treating malignancies of the various organ systems, including but are not limited to, lung, breast, lymphoid, gastrointestinal, and genito-urinary tract adenocarcinomas, and other malignancies such as most colon cancers, renal-cell carcinoma, prostate cancer, non-small cell carcinoma of the lung, cancer of the small intestine, and cancer of the esophagus. A polynucleotide encoding the zinc finger-nucleotide binding polypeptide is also useful in treating non-malignant cell-proliferative diseases such as psoriasis, pemphigus vulgaris, Behcet's syndrome, and lipid histiocytosis. Essentially, any disorder that is etiologically linked to the activation of a zinc finger-nucleotide binding motif containing promoter, structural gene, or RNA, would be considered susceptible to treatment with a polynucleotide encoding a derivative or variant zinc finger derived-nucleotide binding polypeptide.

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLE 1

Construction and Testing of Designed Specific Zinc Finger Domains

Variant zinc finger proteins have been designed and constructed to selectively bind to specific DNA sequences (Table 1). Table 1, below, summarizes the sequences (SEQ ID NO: 77-92) showing the highest selectivity for the sixteen embodiment of GNN target triplets.

Table 1

Target Specificity	Amino acids positions -1 1 2 3 4 5 6	SEQ ID NO:
GAA	Q S S N L V R	77
GAC	D P G N L V R	78
GAG	R S D N L V R	79
GAT	T S G N L V R	80
GCA	Q S G D L R R	81
GCC	D C R D L A R	82
GCG	R S D D L V K	83
GCT	T S G E L V R	84
GGA	Q R A H A L E R	85
GGC	D P G H L V R	86
GGG	R S D K L V R	87
GGT	T S G H L V R	88
GTA	Q S S S L V R	89
GTC	D P G A L V R	90
GTG	R S D E L V R	91
GTT	T S G S L V R	92

Oligonucleotides for zinc finger library panning

Biotinylated, hairpin-structured target site oligos for panning of finger 2 ("F2") libraries had the following sequence:

F2XXX:

25' -Biotin-GGA CGC N'N'N' CGC GGG TTTT CCC GCG NNN GCG TCC-3'
(SEQ ID NO: 25) where NNN = either of the 16 triplets of the GNN set,
or TGA and N'N'N' = its complement.

Non-biotinylated, hairpin structured specific competitor oligos had the following sequence:

F2NNN:

5'-GGA CGC N'N'N' CGC GGG TTTT CCC GCG NNN GCG TCC-3' (SEQ ID NO: 25) where NNN = a mixture of all 64 existing triplets and N'N'N' = its complement.

Panning of zinc finger libraries

Panning of zinc finger phage display libraries was carried out in solution using biotinylated target site hairpin oligos. Seven rounds of panning were carried out as follows: Phage prepared from an overnight culture was allowed to pre-bind to varying amounts of non-biotinylated specific competitor hairpin oligo prior to the addition of the target site oligo. The pre-binding was carried out in 400 μ l Zinc buffer A containing 1% Blotto, 5mM DTT, 4 μ g sheared herring sperm DNA and 100 μ l phage preparation. Typically, 10 times less specific competitor than target oligo was used for the first round of panning. For the subsequent panning rounds, the amount of specific competitor was gradually increased, up to a maximum of 12 μ g in the last panning round(s). Following 30 minutes at room temperature, 100 μ l Zinc buffer A containing 0.4 μ g biotinylated target hairpin oligo were added. After 2.5 to 3.5 hours at RT, phage bound to the target oligo was collected by the addition of 50 μ l Dynabeads M-280 suspension (Dynal) and incubation for one hour at RT. The beads were collected with a magnet, washed 10 times with Zinc buffer A (10 mM Tris, pH 7.5 / 90 mM KCl / 1 mM MgCl₂/ 90 μ M ZnCl₂) containing 2% Tween-20 and 5mM DTT, and once with Zinc buffer A containing 5mM DTT. Phage was eluted for 30 minutes at RT with 25 μ l of TBS containing 10 mg/ml trypsin. Following the addition of 75 μ l Super Broth, eluted phage was allowed to infect 5ml of *E. coli* ER2537 culture for 30 minutes in a 37 degrees Celsius shaker. The volume was increased to 10ml and Carbenicillin was added to a concentration of 20 μ g/ml. At this stage, the number of output phage was determined by plating aliquots of the infected bacteria onto Carbenicillin-containing LB-agar plates. After one hour shaking at 37 degrees Celsius, the Carbenicillin concentration was increased to 50 μ g/ml. After one more hour shaking at 37 degrees Celsius, 10¹³ pfu helper phage was added and the culture was incubated for a few minutes at RT. Then, 90ml of Super Broth containing Carbenicillin (50 μ g/ml) and ZnCl₂ (90 μ M) were added and the culture

was incubated at 37 degrees Celsius for two hours. Upon addition of Kanamycin to a final concentration of 70 µg/ml, the culture was incubated in a 37 degrees Celsius shaker overnight. Phage was purified from culture supernatants by PEG precipitation and resuspended in 2 ml Zinc buffer A containing 1% BSA and 5 mM DTT for further rounds of panning. The number of phage was determined by using various dilutions of the phage prep for infection of *E. coli* ER2537, followed by plating onto Carbenicillin-containing LB-agar plates. Following seven rounds of panning, zinc finger cDNAs were subcloned into the bacterial expression vector pMal-CSS, a derivative of pMal-C2 (New England Biolabs), allowing for expression of the zinc finger proteins as maltose binding protein (MBP) fusions.

Generation of proteins with desired DNA binding specificity.

To generate DNA encoding three-finger proteins, F2 coding regions were PCR amplified from selected or designed F2 variants and assembled by PCR overlap extension. Alternatively, DNAs encoding three-finger proteins with a Zif268 or Sp1C framework were synthesized from 8 or 6 overlapping oligonucleotides, respectively. Sp1C framework constructs were generated as follows.

In the case of E2C-HS1(Sp1), 0.4 pmole each of oligonucleotides SPE2-3 (5'-GCG AGC AAG GTC GCG GCA GTC ACT AAA AGA TTT GCC GCA CTC TGG GCA TTT ATA CGG TTT TTC ACC-3' (SEQ ID NO: 26) and SPE2-4 (5'GTG ACT GCC GCG ACC TTG CTC GCC ATC AAC GCA CTC ATA CTG GCG AGA AGC CAT ACA AAT GTC CAG AAT GTG GC-3') (SEQ ID NO: 27) were mixed with 40 pmole each of oligonucleotides SPE2-2 (5'-GGT AAG TCC TTC TCT CAG AGC TCT CAC CTG GTG CGC CAC CAG CGT ACC CAC ACG GGT GAA AAA CCG TAT AAA TGC CCA GAG-3') (SEQ ID NO: 28) and SPE2-5 (5'-ACG CAC CAG CTT GTC AGA GCG GCT GAA AGA CTT GCC ACA TTC TGG ACA TTT GTA TGG C-3') (SEQ ID NO:29) in a standard PCR mixture and cycled 25 times (30 seconds at 94 degrees Celsius, 30 seconds at 60 degrees

Celsius, 30 seconds at 72 degrees Celsius). An aliquot of this pre-assembly reaction was then amplified with 40 pmole each of the primers SPE2-1 (5'-GAG GAG GAG GAG GTG GCC CAG GCG GCC CTC GAG CCC GGG GAG AAG CCC TAT GCT TGT CCG GAA TGT GGT AAG TCC 5 TTC TCT CAG AGC-3') (SEQ ID NO: 30) and SPE2-6 (5'-GAG GAG GAG GAG CTG GCC GGC CTG GCC ACT AGT TTT TTT ACC GGT GTG AGT ACG TTG GTG ACG CAC CAG CTT GTC AGA GCG-3') (SEQ ID NO: 31) using the same cycling conditions.

The E2C-HS2(Sp1), B3B-HS1(Sp1), B3B-HS2(Sp1), B3C2-
10 HS1(Sp1), and B3C2-HS2(Sp1) DNAs were generated in the same way, using analogous sets of oligonucleotides differing only in the recognition helix coding regions. All assembled three-finger coding regions were digested with the restriction endonuclease *Sfi*1 and cloned into pMal-CSS, a derivative of the bacterial expression vector pMal-C2 (New England
15 Biolabs), allowing for expression of the zinc finger proteins as MBP fusions. DNAs encoding six-finger proteins with each of the different frameworks were assembled in pMal-CSS using *Xma*1 and *Bsr*F1 restriction sites included in the sequences flanking the three-finger coding regions (Beerli *et al.* (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:14628-
20 14633).

Preparation of MBP-zinc finger fusion proteins for ELISA assays

Plasmid pMal constructs containing the zinc-finger coding sequences were transformed into the *E. coli* strain XL1-Blue by electroporation. Three milliliters of Super Broth were inoculated and grown overnight at 37 degrees Celsius. The next day, the cultures were diluted 1:20 in 50 ml conical tubes and grown at 37 degrees Celsius until OD₆₀₀ = 0.5. IPTG was added to a final concentration of 0.3 mM, and incubation was continued for 2 hours. The cultures were centrifuged for 20 minutes, then the pellets resuspended in 400 µl of Zinc Buffer A containing 5 mM fresh DTT. The samples were then frozen in dry ice/ethanol and thawed in 37 degrees Celsius water 6 times, then finally

centrifuged for 30 seconds and left on ice for 30 minutes before use of the supernatants.

ELISA assays

Streptavidin at a concentration of 0.2 μ g/25 μ l in PBS was added to each well of a 96 well plate, then incubated for 1 hour at 37 degrees Celsius. The plate was washed 2x with water, then biotinylated oligo at 0.1 μ g/25 μ l in PBS, or just PBS, was added to the appropriate wells and incubated for 1 hour at 37 degrees Celsius. The plate was washed 2x with water, then each well was filled with 3% BSA in PBS and incubated for 1 hour at 37 degrees Celsius. The BSA was removed without washing, and 25 μ l of the appropriate extract diluted in Zinc buffer A containing 5mM DTT was added to the appropriate wells. The binding reaction was allowed to proceed for 1 hour at room temperature. The plate was washed 8x with water, then α -MBP mAb in Zinc buffer A and 1% BSA was added to the wells followed by incubation for 30 minutes at room temperature. The plate was washed 8x with water, then anti-mouse mAb conjugated to alkaline phosphatase in Zinc buffer A was added, and the plate was incubated for 30 minutes at room temperature. After 8 final washes with water, 25 μ l of alkaline phosphatase substrate and developer was added to each well. Incubation was performed at room temperature, and the OD₄₀₅ of each well was determined at 30 minute and 1 hour time points.

Construction of zinc finger-transcription regulating domain fusion proteins

cDNA encoding amino acids 473 to 530 of the *ets* repressor factor (ERF) repressor domain (ERD) (Sgouras *et al.* (1995) *EMBO J.* 14:4781-4793) was generated from four overlapping oligonucleotides using Taq DNA polymerase; a cDNA encoding amino acids 1 to 97 of the KRAB domain of KOX1 (Margolin *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:4509-4513) was assembled from 6 overlapping oligonucleotides; a cDNA encoding amino acids 1 to 36 of the Mad *sin3* interaction domain (SID) (Ayer *et al.* (1996) *Mol. Cell. Biol.* 16:5772-5781) was assembled

from 3 overlapping oligonucleotides. The coding region for amino acids 413 to 489 of the VP16 transcriptional activation domain (Sadowski *et al.*, (1988) *Nature* 335:563-564) was PCR amplified from pcDNA3/C7-C7-VP16 (Liu *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:5525-5530).
5 The VP64 DNA, encoding a tetrameric repeat of VP16's minimal activation domain, comprising amino acids 437 to 447 (Seipel *et al.* (1992) *EMBO* 13:4961), was generated from two pairs of complementary oligonucleotides. All resulting effector domain-encoding fragments were fused to zinc finger coding regions by standard cloning procedures, such
10 that each resulting construct contained an internal SV40 nuclear localization signal, as well as a C-terminal HA decapeptide tag. Fusion constructs were cloned into pcDNA3 for expression in mammalian cells.

Construction of integrin β3 and erbB-2 luciferase reporter plasmids.

An integrin β3 promoter fragment encompassing nucleotides -584 to -1 (with respect to the ATG codon) was PCR amplified from human genomic DNA, using the primers b3p(Nhe1)-f (5'-GAG GAG GAG GCT AGC GGG ATG TGG TCT TGC CCT CAA CAG GTA GG-3') (SEQ ID NO: 32) and b3p(Hind3)-b (5'-GAG GAG GAG AAG CTT CTC GTC CGC CTC CCG CGG CGC TCC GC-3') (SEQ ID NO: 33), and Taq Expand DNA Polymerase mix (Boehringer). The cycling conditions were: 30 minutes at 94 degrees Celsius; 40 x (one minute at 94 degrees Celsius - 30 minutes at 62 degrees Celsius - 2.5 minutes at 72 degrees Celsius); 10 minutes at 72 degrees Celsius. 10% DMSO was present in the reaction mix.
20

An erbB-2 promoter fragment (Ishii *et al.* (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84:4374-4378) encompassing nucleotides -751 to -1 was PCR amplified under the same conditions, using the primers e2p(Nhe1)-f (5'-GAG GAG GAG GCT AGC CGA TGT GAC TGT CTC CTC CCA AAT TTG TAG ACC-3') (SEQ ID NO: 34) and e2p(Hind3)-b (5'-GAG GAG GAG AAG CTT GGT GCT CAC TGC GGC TCC GGC CCC ATG-3') (SEQ ID NO: 35). PCR products were purified with the Qiagen PCR prep kit, digested
25
30

with the restriction endonucleases Nhe1 and Hind3, and cloned into pGL3basic (Promega).

An *erbB-2* promoter fragment encompassing nucleotides -1571 to -24 was excised from pSVOALΔ5'/erbB-2(N-N) by Hind3 digestion and 5 subcloned into pGL3basic. pSVOALΔ5'/erbB-2(N-N) was a gift from Gordon Gill.

Luciferase assays

For all transfections, HeLa cells were plated in 24 well dishes and used at 10 a confluence of 40-60%. Typically, 200 ng reporter plasmid (pGL3-promoter constructs or, as negative control, pGL3basic) and 20 ng effector plasmid (zinc finger constructs in pcDNA3 or, as negative control, empty pcDNA3) were transfected using the lipofectamine reagent (Gibco BRL). Cell extracts were prepared approximately 48 hours after transfection. Luciferase activity was measured with the Promega 15 luciferase assay reagent, in a MicroLumat LB96P luminometer (EG&G Berthold).

Selection strategy for the generation of six-finger proteins with DNA binding specificity

Based on the modular nature of zinc finger domains, as well as the 20 fact that each zinc finger recognizes 3 bp of DNA sequence, several strategies can be employed to generate zinc finger proteins, with preferably one to three fingers, with desired DNA binding specificity an. For instance, *in vitro* evolution of a six-finger protein binding an 18bp target sequence can follow the strategy outlined in FIGURE 1. The target 25 sequence is divided into six 3bp sub-sites, A-F. In the first step, a Zif268-based zinc finger phage display library in which the central finger 2 is randomized is selected against all 6 subsites in the context of the 2 wild type fingers. After successfull generation of all the finger 2 variants required for a given target, cDNAs encoding three-finger proteins 30 recognizing either half-site 1 (ABC) or half-site 2 (DEF) are constructed via PCR overlap extension. Finally, standard cloning procedures are used to

construct a gene encoding a six-finger protein recognizing the whole 18bp target site.

As an alternative to the serial connection of F2 domain variants, three-and six-finger proteins can be produced by "helix grafting". The framework residues of the zinc finger domains, those residues that support the presentation of the recognition helix, vary between proteins. The framework residues play a role in affinity and specificity. Thus, amino acid positions -2 to 6 of the DNA recognition helices are either grafted into a Zif268 (Pavletich *et al.* (1991) *Science* 252:809-817) or an Sp1C framework (Desjarlais *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:2256-2260).

Choice of human integrin β_3 and erbB-2 target sequences

Panning experiments carried out previously indicated that zinc fingers binding to G-containing triplets, with a G or a T in 5'-position, are more readily obtained than zinc fingers binding other triplets. The zinc finger target sequences were selected such that they contained one or more G's in each triplet of the 18bp sequence, and that each triplet started with a G or a T (Table 2). To conform with these requirements, erbB-2 target B2 was split into two halves separated by two bases. A longer linker peptide between the appropriate zinc fingers may also for recognition of such a split site. Blast sequence similarity searches were carried out with each of the target sequences and confirmed that each 18bp sequence specifies a unique site in the human genome (maximal similarity tolerated: 16/18bp identity).

Since transcription factor AP-2 is involved in deregulated expression of erbB-2 in a significant fraction of ErbB-2 overexpressing tumor cell lines, erbB-2 target site B2 was designed to overlap with the AP-2 binding site GCTGCAGGC, with the intention of inhibiting expression of ErbB-2 not only as a result of active transcriptional repression, but also by competition with an important transcription factor. In contrast, zinc finger proteins binding the other erbB-2 target sites (i.e.

erbB-2 target sites C and D), affect transcription as a result of their effector domains.

Integrin $\beta 3$ target sequences B and C2 were chosen at various distances from the transcription start site, to allow for a comparison of the efficacy of transcriptional regulation. Since the selected zinc finger proteins are fused to transcriptional effector domains (Sadowski *et al.*, 5 (1988) *Nature* 335:563-564; Margolin *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:4509-4513) ; Sgouras *et al.* (1995) *EMBO* 14:4781-4793; Ayer *et al.* (1996) *Mol. Cell. Biol.* 16:5772-5781), binding of a zinc finger protein *per se* have an effect on the level of transcription.

A list of chosen target sequences for the selection of zinc finger proteins is given in Table 2, below. Since zinc finger proteins make base contacts predominantly with one strand of the DNA double helix, only the relevant strand of the target sequence is listed and designated +/- with respect to the coding strand. The location of the target sequences and position relative to the major transcription start site(s) is given.

Table 2: Chosen Target Sequences For The Selection Of Zinc Finger Proteins

Integrin $\beta 3$ (B3) target sequences		LOCATION	SEQ ID
B3B	GCC TGA GAG GGA GCG GTG	- strand, promoter region, -160bp	72
B3C2	GGA GGG GAC GCG GTG GGT	- strand, promoter region, -70bp	73
ErbB-2 (E2) target sequences			
E2B2	GTG TGA GAA(CG)GCT GCA GGC	+ strand, promoter, -150/-220bp	74
E2B2	GTG TGA GAA(CG)GCT GCA GGC	+ strand, promoter, -150/-220bp	74
E2C	GGG GCC GGA GCC GCA GTG	+ strand, 5' UTR, +160/+230bp	75
E2D	GCA GTT GGA GGG GGC GAG	+ strand, promoter, -30/-100bp	76

Construction and panning of a finger 2 library

The amino acid residues implicated in contacting DNA in finger 2 of the Zif268-C7 ("C7") (Wu *et al.* (1995) *Proc. Natl. Acad. Sci. U.S.A.* 30 92:344-348) have been extensively randomized using the PCR overlap extension mutagenesis strategy. Using two different randomization strategies, two sublibraries have been constructed using the pComb3

phage display vector (Barbas *et al.* (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88:7978-7982). The sublibraries contain approximately 4×10^9 independent clones each.

The mutagenesis strategy for randomization of finger 2 of Zif268-C7, showing helix positions -3 to 7, is summarized in Table 3, below. The top line shows the wild type sequence of finger 2. The lower two lines show the two mutagenesis strategies used, where N = G, A, T, C; K = G, T; V = G, A, C; S = G, C. The NNK randomized codon provides all 20 amino acids in 32 codons. The VNS randomized codon provides 16 amino acids in 24 codons, excluding Phe, Trp, Tyr, Cys and all stops. Note that in the strategy shown in the bottom line, the use of less complex codons allows for the mutagenesis of an additional codon.

Table 3:
Mutagenesis strategy for randomization of finger 2
of Zif268-C7, showing helix positions -3 to 7.

-3	-2	-1	1	2	3	4	5	6	7
F	S	R	S	D	H	L	T	T	H
F	S	(NNK)	(NNK)	(NNK)	(NNK)	L	(NNK)	(NNK)	H
F	(VNS)	(VNS)	(VNS)	(VNS)	(VNS)	L	(VNS)	(VNS)	H

Finger 2 variants recognizing each of the 16 triplets of the GXX set (Segal *et al.* (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96:2758-2763; and Table 1), as well as one variant recognizing TGA, have been successfully selected. In extension of previous observations, comparison of the zinc finger sequences revealed a code for zinc finger recognition of DNA. Thus, a 5'- and 3'-G selected an arginine at helix positions 6 and -1, respectively, while a central G selected an a histidine or lysine at position 3 of the recognition helix. In contrast, a central A selected an asparagine, a 3'-A a glutamine, a central T a serine or alanine, a 3'-T a threonine or serine, a central C an aspartate or threonine, and a 3'-C an aspartate or glutamate at the corresponding helix positions. An extensive characterization of the specificities and affinities of selected zinc finger variants has been carried out and indicates that many of the zinc finger

peptides recognize their targets in a highly specific manner (Segal *et al.* (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96:2758-2763 and Table 1).

Refinement of finger 2 specificities by site-directed mutagenesis

Attempts were made to improve binding specificity of some of the zinc finger domains by modifying the recognition helices by using site-directed mutagenesis. Data from the phage display selections and structural information guided the design of the mutants. Although helix positions 1 and 5 were not expected to play a direct role in recognition, the best improvements in specificity always involved modifications in these positions (Segal *et al.* (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96:2758-2763 and Table 1). These residues have been observed to make phosphate backbone contacts, which contribute to affinity in a nonsequence-specific manner. Thus, removal of nonspecific contacts can increase the importance of the specific contacts to the overall stability of the complex, thereby enhancing specificity.

Generation of three finger proteins binding erbB-2 and integrin β 3 target sequences

Two different strategies for generating three-finger proteins recognizing 9 bp of DNA sequence were used. Each strategy is based on the modular nature of the zinc finger domain, and takes advantage of a family of zinc finger domains recognizing triplets of the 5'-GNN-3' type defined in Table 1. Two three-finger proteins recognizing half sites (HS) 1 and 2 of the 5'-(GNN)₆-3' erbB-2 target site e2c were generated in the first strategy by fusing the pre-defined finger 2 (F2) domain variants together using a PCR assembly strategy.

To examine the generality of this approach, three additional three-finger proteins recognizing sequences of the 5'-(GNN)₃-3' type, were prepared using the same approach. Purified zinc finger proteins were prepared as fusions with the maltose binding protein (MBP). ELISA analysis revealed that serially connected F2 proteins were able to act in concert to specifically recognize the desired 9-bp DNA target sequences

(Beerli *et al.* (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:14628-14633). Each of the 5 proteins shown was able to discriminate between target and non-target 5'-(GNN)₃-3' sequence. The affinity of each of the proteins for its target was determined by electrophoretic mobility-shift assays. These studies demonstrated that the zinc finger peptides have affinities comparable to Zif268 and other natural transcription factors with K_d values that ranged from 3 to 70 nM (Table 4, below).

As an alternative to the serial connection of F2 domain variants, in the second strategy, three-finger proteins specific for the two halfsites of the erbB-2 target site e2c (Table 4, below), were produced by "helix grafting." The framework residues may play a role in affinity and specificity. For helix grafting, amino acid positions -2 to 6 of the DNA recognition helices were either grafted into a Zif268 (Pavletich *et al.* (1991) *Science* 252:809-817) or an Sp1C framework (Desjarlais *et al.* (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90:2256-2260). The Sp1C protein is a designed consensus protein shown to have enhanced stability towards chelating agents. The proteins were expressed from DNA templates prepared by a rapid PCR-based gene assembly strategy. In each case, ELISA analysis of MBP fusion proteins showed that the DNA binding specificities and affinities (Table 4, below) observed with the F2 framework constructs were retained. Three finger proteins recognizing HS1 and HS2 of the integrin β 3 target sites b3b and b3c2 have also been generated, using the Sp1C backbone. Preliminary ELISA data showed that these proteins bind their respective targets with good specificity. Further characterization of proteins can be made, such as determination of their affinities by gel shift analysis. See Table 4, below.

Generation of six-finger proteins for specific targeting of the erbB-2 and integrin β 3 promoter regions.

The recognition of 9 bp of DNA sequence is not sufficient to specify a unique site within a complex genome. In contrast, a six-finger protein recognizing 18 bp of contiguous DNA sequence could define a

single site in the human genome, thus fulfilling an important prerequisite for the generation of a gene-specific transcriptional switch. Six-finger proteins binding the *erbB-2* target sequence e2c were generated from three-finger constructs by simple restriction enzyme digestion and cloning with F2, Zif268, and Sp1C framework template DNAs (for sequences of these proteins, see Beerli *et al.* (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:14628-14633). Six finger proteins binding the integrin $\beta 3$ target sequences b3b and b3c2 were only generated using the Sp1C backbone. ELISA analysis of purified MBP fusion proteins showed that each of the six-finger proteins was able to recognize the specific target sequence, with little cross reactivity to non-target 5'-(GNN)₆-3' sites or a tandem repeat of the Zif268 target site.

In Table 4, below, the affinities of three- and six-finger proteins for various target sequences as determined by gel shift analysis is summarized. Proteins are named with upper case letters, DNA target sequences with lower case letters. Abbreviations used are: F2 = finger 2 framework; Zif = Zif268 framework; Sp1 = Sp1C framework; mut = mutant; HS = half-site. With respect to the target site overlap phenomenon, the base following each target sequence is given in lower case letter (see Beerli *et al.* (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:14628-14633). The affinity of the Zif268-DNA interaction was determined to be 10 nM (Segal *et al.* (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96:2758-2763). K_d values are averages from 2 independent experiments, with standard deviations of 50% or less.

Table 4:
AFFINITIES OF THREE AND SIX FINGER PROTEINS

	Protein	Target	Target Sequence (5'-3')	K_d , nM
5	B3(F2)	b3	GGA GGG GAC g	4
	E2(F2)	e2	GGG GGC GAG g	3
	C5(F2)	c5	GGA GGC GGG g	30
10	E2C-HS1(F2)	e2c-hs1	GGG GCC GGA g	45
	E2C-HS1(Zif)	e2c-hs1	GGG GCC GGA g	70
	E2C-HS1(Sp1)	e2c-hs1	GGG GCC GGA g	35
15	E2C-HS2(F2)	e2c-hs2	GCC GCA GTG g	70
	E2C-HS2(Zif)	e2c-hs2	GCC GCA GTG g	75
	E2C-HS2(Sp1)	e2c-hs2	GCC GCA GTG g	25
20	E2C(F2)	e2c-g	GGG GCC GGA GCC GCA GTG g	25
	E2C(Zif)	e2c-g	GGG GCC GGA GCC GCA GTG g	1.6
	E2C(Zif)	e2c-a	GGG GCC GGA GCC GCA GTG a	2.3
25	E2C(Zif)	e2c-muths1	AGT CTG AAT GCC GCA GTG g	200
	E2C(Zif)	e2c-muths2	GGG GCC GGA AGT CTG AAT g	200
	E2C(Sp1)	e2c-g	GGG GCC GGA GCC GCA GTG g	0.5
30	E2C(Sp1)	e2c-a	GGG GCC GGA GCC GCA GTG a	0.75
	E2C(Sp1)	e2c-muths1	AGT CTG AAT GCC GCA GTG g	65
	E2C(Sp1)	e2c-muths2	GGG GCC GGA AGT CTG AAT g	100

In Table 5, below, the finger 2 variants generated by phage display selection and refined by site-directed mutagenesis are summarized. Protein designations are in the form pXXX, for clones derived from panning; pmXXX refers to clones refined by mutagenesis. Helix positions -1, 3, and 6 are shown in **bold**, altered nucleotides are underlined. The values represent the results of at least two independent experiments. The standard error was $\pm 50\%$ or less.

Table 5:
SUMMARY OF FINGER 2 VARIANTS GENERATED BY PHAGE DISPLAY
SELECTION AND REFINED BY SITE-DIRECTED MUTAGENESIS

	Protein	Finger-2 Helix	Finger-2 Subsite	K _D (nM)	K _{D, Prot} /K _{D, Zif268}
5	pGGG	SRSDHLTR	GGG	0.4	0.04
	pmGGG	SRSDKLVR	GGG	6	0.6
	"	"	<u>GTG</u>	>1,400	
	pGGA	SQRAHLER	GGA	3	0.3
10	pmGGT	STSGHLVR	GGT	15	1.5
	"	"	<u>GGC</u>	>2,400	
	pmGGC	SDPGHLVR	GGC	40	4.0
	pmGAG	SRSDNLVR	GAG	1	0.1
	"	"	<u>GCG</u>	45	4.5
15	pmGAA	SQSSNLVR	GAA	0.5	0.05
	pGAT	STSGNLVR	GAT	3	0.3
	pmGAC	SDPGNLVR	GAC	3	0.3
	"	"	<u>GCC</u>	90	9.0
	pGTG	SRKDSLVR	GTG	3	0.3
20	pmGTG	SRSDELVR	GTG	15	1.5
	"	"	<u>GAG</u>	30	3.0
	pGTA	SQSSSLVR	GTA	25	2.5
	"	"	<u>GTG</u>	>1,000	
	pmGTT	STSGSLVR	GTT	5	0.5
	pGTC	SDPGALVR	GTC	40	4.0
25	"	"	<u>GCC</u>	>4,400	
	pmGCG	SRSDDLVR	GCG	9	0.9
	"	"	<u>GAG</u>	6	0.6
	pGCA	SQSGDLRR	GCA	2	0.2
	"	"	<u>GCT</u>	10	1
30	pmGCT	STSGELVR	GCT	65	6.5
	pGCC	SDCRDLAR	GCC	80	8.0
	pTGA	SQAGHLAS	TGA	nd	nd
	C7	SRSDHLTT	TGG	0.5	0.05
	Zif268	SRSDHLTT	TGG	10	1

35

The affinity of each of the E2C proteins for the e2c DNA target site was determined by gel-shift analysis. A modest K_d value of 25 nM was observed with the E2C(F2) six-finger protein constructed from the F2

framework (Table 5, above; Beerli *et al.* (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:14628-14633), a value that is only 2 to 3 times better than its constituent three-finger proteins. In previous studies of six-finger proteins, an approximately 70-fold enhanced affinity of the six-finger proteins for their DNA ligand compared to their three-finger constituents was observed (Liu *et al.* (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:5525-5530). The absence of a substantial increase in the affinity of the E2C(F2) peptide suggested that serial connection of F2 domains is not optimal. It is possible that the periodicity of the F2 domains of the six-finger protein does not match that of the DNA over this extended sequence, and that a significant fraction of the binding energy of this protein is spent in unwinding DNA. In contrast to the F2 domain protein, the E2C(Zif) and E2C(Sp1) six-finger proteins displayed 40- to 70-fold increased affinity as compared to their original three-finger protein constituents, with K_d values of 1.6nM and 0.5nM, respectively. Significantly, both three-finger components of these proteins were involved in binding, since mutation of either half-site led to a roughly 100-fold decrease in affinity (Table 4, above; Beerli *et al.* (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:14628-14633). The preponderance of known transcription factors bind their specific DNA ligands with nanomolar affinity, suggesting that the control of gene expression is governed by protein/DNA complexes of unexceptional life times. Thus, zinc finger proteins of increased affinity should not be required and could be disadvantageous, especially if binding to non-specific DNA is also increased. The affinities of the B3B(Sp1) and B3C2(Sp1) six finger proteins for their respective targets can be determined by one skilled in the art using well-known methods as well as those described herein.

EXAMPLE 2

Construction Of Fusion Proteins Containing Zinc Finger Domains and Transcriptional Repressors And Activators

In order to demonstrate use of zinc finger proteins as gene-specific transcriptional regulators, the E2C(Sp1), B3B(Sp1), and B3C2(Sp1) six-finger proteins were fused to a number of effector domains (Beerli *et al.* (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:14628-14633). Transcriptional repressors were generated by attaching either of three human-derived repressor domains to the zinc finger protein. The first repressor protein was prepared using the ERF repressor domain (ERD) (Sgouras *et al.* (1995) *EMBO J.* 14:4781-4793), defined by amino acids 473 to 530 of the ets2 repressor factor (ERF). This domain mediates the antagonistic effect of ERF on the activity of transcription factors of the *ets* family. A synthetic repressor was constructed by fusion of this domain to the C-terminus of the zinc finger protein.

The second repressor protein was prepared using the Krüppel-associated box (KRAB) domain (Margolin *et al.* (1994) *Proc. Natl. Acad. Sci. USA* 91:4509-4513). This repressor domain is commonly found at the N-terminus of zinc finger proteins and presumably exerts its repressive activity on TATA-dependent transcription in a distance- and orientation-independent manner, by interacting with the RING finger protein KAP-1. The KRAB domain found between amino acids 1 and 97 of the zinc finger protein KOX1 was used. In this case an N-terminal fusion with the six-finger protein was constructed. Finally, to demonstrate the utility of histone deacetylation for repression, amino acids 1 to 36 of the Mad mSIN3 interaction domain (SID) were fused to the N-terminus of the zinc finger protein (Ayer *et al.*, (1996) *Mol. Cell. Biol.* 16:5772-5781). This small domain is found at the N-terminus of the transcription factor Mad and is responsible for mediating its transcriptional repression by interacting with mSIN3, which in turn interacts the co-repressor N-CoR and with the histone deacetylase mRPD1.

To examine gene-specific activation, transcriptional activators were generated by fusing the zinc finger protein to amino acids 413 to 489 of the herpes simplex virus VP16 protein (Sadowski *et al.* (1988) *Nature* 335:563-564), or to an artificial tetrameric repeat of VP16's minimal activation domain, DALDDFDLDML (SEQ ID NO: 36) (Seipel *et al.* (1992) *EMBO* 13:4961), designated VP64.

Specific regulation of erbB-2 promoter activity

Reporter constructs containing fragments of the *erbB-2* promoter coupled to a luciferase reporter gene were generated to test the specific activities of the *erbB-2* specific synthetic transcriptional regulators. The target reporter plasmid contained nucleotides -758 to -1 with respect to the ATG initiation codon, whereas the control reporter plasmid contained nucleotides -1571 to -24, thus lacking all but one nucleotide of the E2C binding site encompassed in positions -24 to -7. Both promoter fragments displayed similar activities when transfected transiently into HeLa cells, in agreement with previous observations. To test the effect of zinc finger-repressor domain fusion constructs on *erbB-2* promoter activity, HeLa cells were transiently co-transfected with each of the zinc finger expression vectors and the luciferase reporter constructs (Beerli *et al.*, (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:14628-14633). Significant repression was observed with each construct. The ERD and SID fusion proteins produced approximately 50% and 80% repression, respectively. The most potent repressor was the KRAB fusion protein. This protein caused complete repression of *erbB-2* promoter activity. The observed residual activity was at the background level of the promoter-less pGL3 reporter. In contrast, none of the proteins caused significant repression of the control *erbB-2* reporter construct lacking the E2C target site, demonstrating that repression is indeed mediated by specific binding of the E2C(Sp1) protein to its target site. Expression of a zinc finger protein lacking any effector domain resulted in weak repression, approximately 30%, indicating that most of the repression observed with the SID and

KRAB constructs is caused by their effector domains, rather than by DNA-binding alone. This observation strongly suggests that the mechanism of repression is active inhibition of transcription initiation rather than of elongation. Once initiation of transcription by RNA polymerase II has occurred, the zinc finger protein appears to be readily displaced from the DNA by the action of the polymerase.

The use of *erbB-2* specific zinc finger proteins to mediate activation of transcription was demonstrated using the same two reporter constructs. The VP16 fusion protein was found to stimulate transcription approximately 5-fold, whereas the VP64 fusion protein produced a 27-fold activation. This dramatic stimulation of promoter activity caused by a single VP16-based transcriptional activator is exceptional in view of the fact that the zinc finger protein binds in the transcribed region of the gene. This again demonstrates that mere binding of a zinc finger protein, even with one with sub-nanomolar affinity, in the path of RNA polymerase II need not necessarily negatively affect gene expression.

Based on the efficient and specific regulation of a reporter construct driven by the *erbB-2* promoter, the effect of transiently transfected zinc finger expression plasmids on activity of the endogenous *erbB-2* promoter was analyzed. As a read-out of *erbB-2* promoter activity, ErbB-2 protein levels were analyzed by Western blotting. Significantly, E2C(Sp1)-VP64 lead to an upregulation of ErbB-2 protein levels, while E2C(Sp1)-SKD lead to its downregulation. This regulation was specific, since no effect was observed on expression of EGFR.

It is important to note that the observations made in these experiments drastically underestimate the efficacy of the zinc finger peptides, since the transfection efficiency of HeLa cells is no more than 50%. To ascertain that 100% of the cells express the zinc finger proteins stable cell lines need to be generated. Production of stable cell lines expressing the zinc finger constructs under control of a tetracycline-

inducible promoter is known (Gossen *et al.* (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89:5547-5551). Inducible expression of zinc finger proteins in stable cell lines allows for detailed analysis of the degree of specificity of such proteins.

5 *Specific regulation of integrin β3 promoter activity*

To test the activity of transcriptional regulators specific for the integrin β3 promoter, a reporter plasmid was constructed containing the luciferase open reading frame under control of the integrin β3 promoter. When compared to the two erbB-2 promoter fragments described above, 10 the integrin β3 promoter fragment had a very low activity. In fact, in some experiments no activation of luciferase expression over background was detected, preventing an analysis of the effects of the KRAB fusion proteins. However, when the VP64 fusion proteins were tested an efficient activation of the integrin β3 promoter was observed. B3B(Sp1)-VP64 and B3C2(Sp1)-VP64 stimulated transcription 12 and 22fold, 15 respectively. Activation of transcription was specific, since no effect on the activity of the erbB-2 promoter was detected.

EXAMPLE 3

Fusion Protein Construct Comprising Progesterone Receptor Variant

20 Amino acid sequence comparisons of steroid receptor family members indicates that they generally comprise a number of defined domains, including an N-terminal DNA binding domain and a more C-terminally located ligand binding domain. Importantly, these domains are modular and the DNA binding domain of progesterone receptor (PR) has been successfully exchanged for the Gal4 DNA binding domain. The 25 addition of a VP16 activation or a KRAB repressor domain to the N- or C-terminus of this construct yielded proteins that could regulate a Gal4 responsive reporter in a ligand dependent manner. An important feature of the ligand binding domain used in these studies is that it is derived from a mutant PR with a small C-terminal deletion. This mutant fails to respond 30

to progesterone and is responsive only to progesterone antagonists such as RU486, making this system suitable for *in vivo* applications.

The original PR DNA binding domain can be replaced by engineered zinc finger proteins. For example, the three finger protein Zif268(C7) was fused to the N-terminus of the PR ligand binding domain (PBD) (aa 640 to 914), and the VP16 activation domain to its C-terminus. It was found that this fusion protein protein was able to regulate an SV40 promoter luciferase construct with ten upstream Zif268(C7) binding sites in an RU486-dependent manner.

An RU486 dose response curve showed that optimal induction occurs at about 1nM to about 10nM RU486. A time course study was carried out with 10nM RU486 and showed that optimal induction of C7-PBD-VP16 activity occurs at about 24 hours.

Since naturally occurring steroid receptors bind DNA as dimers, an important prerequisite for the application of this approach is the presence of suitable target sequences in the promoter of interest. Fortunately, the spacing and orientation of the two half-sites targeted by steroid receptor dimers is flexible. While a steroid response element usually includes an inverted repeat, or palindrome, also direct repeats or even everted repeats of the half-sites in variable spacing are tolerated (Aumais *et al.* (1996) *J. Biol. Chem.* 272:12229-12235). A search of the *erbB-2* and integrin $\beta 3$ promoters revealed that direct and inverted repeats of 5'-(GNN)₃-3' sequence motifs occur quite frequently. An example of a sequence motif suitable for targeting by a heterodimeric RU486-regulatable zinc finger protein is 5' **GAG GAG GGC TGCTT GAG GAA GTA-3'** (SEQ ID NO: 37), which was found in the *erbB-2* promoter and overlaps with the TATA box (underlined above). In some instances, promoter targeting is possible using a homodimer, for example by targeting the sequence 5'-GCC GGA **GCC ATGGG GCC GGA GCC-3'** (SEQ ID NO: 38), which is also found in the *erbB-2* promoter and overlaps with the target sequence e2c (underlined).

EXAMPLE 4

Recombinant Ligand Activated Transcriptional Regulator Fusion Proteins Containing Human Estrogen Receptor Ligand Binding Domains

The human estrogen receptor is shown in FIG. 2 as an example of a steroid receptor protein. The numbers below the rectangle indicate the position of the amino acid residues defining the borders of each domain. A/B is the domain of the amino terminus activation function 1 (AF-1), C is the DNA binding domain, D is called the hinge region, E is the ligand binding domain, which also contains the activation function 2 (AF-2) and F is the portion closest the carboxyl terminal, a domain whose function has not been fully established. The regions of the protein that participate and stabilize the homodimerized complex are distributed in the C, D and E domains. Regions throughout the steroid receptor ligand binding domain (region E in FIG. 2) as well as regions in the native DBD and hinge region (regions C and D respectively) contribute to homodimerization of the receptor. To demonstrate the importance of these regions to the function of the C2H2-containing receptors, proteins containing three different length LBD fragments were constructed. These differing length LBD constructs are designated A, B, and C (FIG. 3). LBD fragment A represents what is generally referred to as the "minimal" LBD fragment. Some studies have suggested the hinge region plays an important role in steroid receptor LBD — chimeric proteins; fragment B represents the LBD plus hinge. The native C or DNA binding region of estrogen receptor contains two zinc fingers of the C4-C4 class. The 5' or amino terminus finger contributes to DNA specific contacts; the 3' finger contributes to stabilizing the DNA binding domain dimer complex. To take advantage of this contribution of the 3' native zinc finger, LBD fragment C, where the 3' native zinc finger is retained and fused directly to the C2H2 zinc finger array, was included.

In order to optimize the ability of the fusion proteins to regulate gene expression, it may be necessary to add additional heterologous

transactivating domains to the receptor. To facilitate these studies, fusion proteins were constructed either with the full length LBD extending to estrogen receptor residue 595, or with LBD fragments truncated at amino acid (aa) 554 to remove the F region. The full-length constructs are
5 referred to as long (L), the truncated versions as short (S). All constructs contain a heterologous transactivation domain (TA) comprised of a VP16 minimal domain, unless otherwise noted, fused to the carboxy terminus of the ligand binding domain. VP16 minimal domain trimer has the amino acid residue sequence 3 x (PADALDDFDLDML) (SEQ ID NO: 36), and is
10 the tetracycline controlled transactivator (tTA) TA2 (Baron *et al.* (1997) *Nucleic Acids Research* 25:2723-2729).

These constructs are summarized in FIG. 3, which provides a schematic summary of the cloning strategy and nomenclature related to the C2H2 DNA binding domain - ER ligand binding domain fusion proteins.
15 As shown in the plasmid construct at the bottom, the final construct contains three components: a C2H2 zinc finger domain (ZFP) at the amino end, a steroid receptor ligand binding domain (LBD) fragment in the middle, and a heterologous transactivation domain (TA) appended onto the carboxyl end. LBD fragments A, B, or C were defined by the position of the amino terminus border of the LBD; amino acid number for A (283),
20 B (258) and C (212) correspond to the residue numbers in wild type ER. LBD fragments were further defined as long (L) or short (S) depending on their carboxy terminus junction. Long constructs fuse the heterologous TA to the wt ER amino acid residue 595, short constructs fuse TA to an LBD fragment truncated at ER amino acid 554. Thus, six fusion proteins in all were constructed, ZFP-LBD-TA A, B and C, each in a long and short form. Maps of specific examples constructed in the expression vector
25 pcDNA3.1 are shown in FIG. 4 (C7LBDAS) (SEQ ID NO: 6), FIG. 5 (C7LBDBS) (SEQ ID NO: 8), FIG. 6 (C7LBDCS) (SEQ ID NO: 10), FIG. 7 (C7LBDAL) (SEQ ID NO: 7), FIG. 8 (C7LBDBL) (SEQ ID NO: 1), and FIG. 9 (C7LBDCL) (SEQ ID NO: 9).

As discussed in detail above, zinc fingers of the C2H2 class each contribute to about 3 bp of DNA sequence contacts. C2H2 zinc finger arrays can be "stitched together" to assemble DNA binding domains having 6, 9, 12, 15, 18 bp or more of specific sequence to which they bind. In order to evaluate the size of the zinc finger array that can be used in these C2H2 Zn finger (ZFP)—steroid receptor fusion proteins, proteins containing 3 finger and 6 finger arrays were constructed. The composition of the various proteins assembled, and their DNA binding site specificity is listed in FIG. 16.

The general cloning strategy was as follows. Three fragments (A, B, and C with reference to FIG. 3) of human estrogen receptor ligand binding domain (LBD) with or without the F region were built into the pcDNA3.1 (Invitrogen) vector backbone through a series of PCR amplification and cloning steps. Initially the LBD fragment A without F region (i.e. short form; LBDAS) and with F region (i.e. long form; LBDAL) were PCR amplified from a plasmid clone of the human wild type estrogen receptor, pHEGO (Tora *et al.* *EMBO J.* 8:1981-1986) with primer pairs NR1/NR2 and NR1/NR3 respectively (Table 1). Convenient restriction sites were incorporated into primers (Table 1) as needed. The PCR amplified LBDAS and LBDAL fragments were first cloned into the SrfI site of pCR-ScriptAmpSK(+) vector (Stratagene), resulting in constructs pLBDAS and pLBDAL. The VP16 minimal domain trimer (TA2; Baron *et al.* (1997) *Nucleic Acids Research* 25:2723-2729) was PCR amplified from plasmid pTTA2 (Clontech) with primer pairs NR4 and NR9 and cloned into the SphI and NotI site of pLBDAS and pLBDAL to generate pLBDASTA2 and pLBDALTA2. To generate LBD fragment B without the F region (LBDBS) and LBD fragment C without the F region (LBDCS), PCR primers NR7 and NR8, which represent the 5' boundary of the LBD region fragment in chimerics B and C respectively were designed (Table 6, below). These primers were paired with the 3' end primer NR6, which incorporates a unique BpuI site in ER. PCR fragments from pHEGO with

primer pair NR6/NR7 and PCR fragment with NR6/NR8 were then cloned into the SpeI and BpI site of pLBDC7ASTA2 backbone. This resulted in plasmid pLBDBSTA2 and pLBDCSTA2.

5

Table 6
PCR Primers Used For Cloning

	NAME / (SEQ ID NO:)	SEQUENCE
10	NR1 (39)	cct act gcc ggc act agt tct gct gga gac atg aga gct gcc aac ctt
	NR2 (40)	cct aaa cgt acg gct agt ggg cgc atg tag gcg gtg ggc gtc
	NR3 (41)	cct aaa cgt acg gac tgt ggc agg gaa acc ctc tgc ctc
	NR4 (42)	cca ctt aaa tgt gaa agt cgt acg ccg gcc
	NR6 (43)	tat ggg ggg ctc agc atc caa caa ggc act
15	NR7 (44)	cct act act agt gac cga aga gga ggg aga atg ttg aaa cac aag cgc
	NR8 (45)	cct act act agt agt att caa gga cat aac gac tat atg tgt
	NR9 (46)	tat cat gtg cgg ccg ctt act tag tta ccc cgg cag cat

Having completed cloning of the three LBD fragments fused to the TA2 region, the C2H2 DNA binding protein C7 was then excised from pcDNAC7VP16 by BglIII and SpeI digestion and ligated into the BamHI and SpeI site of each of the 3 constructions (pLBDASTA2, pLBDBSTA2 and pLBDCSTA2), which resulted in pC7LBDASTA2, pC7LBDBSTA2 and pC7LBDCSTA2. Cassettes of C7LBDASTA2, C7LBDBSTA2 and C7LBDCSTA2 were then removed from the pCR-Script vector by EcoRI-NotI digestion and cloned into the same sites of the expression cassette vector pcDNA3.1(+), resulting in constructs pCDNAC7ASTA2, pCDNAC7BSTA2 and CDNAC7CSTA2. In order to reconstruct these three ZFP-LBD fusion proteins with an LBD fragment including the estrogen receptor F region fused to TA2, the BpI to NotI fragment was excised from pLBDALTA2 construct and substituted for the BpI-NotI fragment in pCDNAC7LBDASTA2, pCDNAC7LBDBSTA2 and pCDNAC7LBDCSTA2 to generate pCDNAC7LBDALTA2, pCDNAC7LBDBLTA2 and pCDNAC7LBDCLTA2.

Cloning for Replacement of DNA Binding Domain C7 with E2C

An intermediate construct pcDNAE2CVP16 was first constructed by replacing the Sfil fragment containing C7 in pcDNAC7VP16 with the E2C(hs1) fragment isolated from pMal/E2C(hs1) after Sfil digestion.

5 Next, pcDNAE2CVP16 was digested with Spel and a 1 kb fragment was isolated. This Spel fragment was ligated to the large Spel fragment of pcDNAC7LBDASTA2, which created pcDNA-E2CLBDASTA2. Similar steps were performed to construct pcDNAE2CLDBSTA2.

Analysis of Recombinant Construct Protein Binding to DNA

10 In order to demonstrate that the fusion proteins bind to DNA in a sequence specific manner, and to evaluate the stoichiometry of protein:DNA binding, standard electrophoretic mobility shift or gel retardation assays were performed.

15 First, fusion proteins were produced by *in vitro* transcription and translation using the TNT Coupled Reticulocyte Lysate System (Promega, Cat # L4610) according to the manufacturer's instructions. Briefly, each expression reaction was set up in a total volume of 50 μ l which contained 25 μ l of TNT rabbit reticulocyte lysate, 2 μ l of TNT Reaction Buffer, 2 μ l of RNasin ribonuclease inhibitor (20 U/ μ l), 1 μ l each of amino acid mixture minus leucine, amino acid mixture minus methionine and TNT T7 RNA polymerase, 2 μ l of expression plasmid (1 μ g/ μ l) and water. The reaction mixture was incubated at 30° C for 90 minutes.

20 Binding of the expressed protein to duplex oligonucleotides was performed as follows, using the gel shift assay systems (Promega, Cat # E3050): 5 μ l of *in vitro* translation product was co-incubated with 4 μ l of 5X gel shift binding buffer and 7 μ l of water at room temperature for 20 min, then 2 μ l of E2 (10 nM final concentration) and 2 μ l of 32 P-labeled probe were added to the mixture. The probe had been labeled using standard protocol as described in the kit. After incubated at room 25 temperature for about 20 minutes, the mixture was loaded onto a 6% 30

DNA retardation gel and run in 0.5X TBE buffer at 150-200 volts for about 30-60 minutes. The gel was then dried and exposed to X-ray film.

A DNA oligonucleotide containing two inverted binding sites for the C2H2 domain known as C7, each half site separated by 3 bp, was used for the initial assessment of DNA binding. This palindromic configuration mimics the composition of the native estrogen receptor response element (ERE), except that the natural 6 bp half site of ERE is replaced by the 9 bp half site specified by C7. Binding of the C7-LBD fusion proteins A, B, and C, all in the short form, were tested and compared to the control proteins C7VP16 and 2C7VP16 (see, Liu, et al. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94:5525-5530, which describes the control proteins). For each protein, binding was tested in the absence or presence of 100 fold excess of unlabeled oligonucleotide (1.75 μ M) as a competitor. Competition of the gel shift product by the unlabeled oligonucleotide indicates the band is a specific protein:DNA interaction. The results demonstrated that C7VP16 can bind once or twice to the oligonucleotide, creating two specific gel shift bands. 2C7VP16 binds only once to the oligonucleotide containing two inverted C7 sites. Notably, C7LBDA and C7LBDB bind strongly to yield one major species, which runs higher than any of the control bands. Although true molecular mass cannot be determined from this type of mobility assay, the relative size of the complexes suggest the protein bound for C7LBD is larger than for C7VP or 2C7VP. The size of the band and presence of only one major species indicate that the fusion protein ZFP-LBD is binding to the oligonucleotide as a dimer. No significant gel shift product was detected for C7LBD chimeric C, suggesting that the addition of the additional native zinc finger from the estrogen receptor may have reduced the affinity of the fusion protein for its C2H2-specific DNA binding site. Finally, the reduction of binding for each of the gel shift products by the addition of the unlabeled oligonucleotide indicates that these fusion proteins are binding to DNA in a sequence specific manner.

To further demonstrate that the chimera ZFP-LBD binds to DNA as a dimer, the binding of C7LBD A, B, and C to oligonucleotides containing one or two C7 binding sites was tested. Three fusion proteins (C7LBDAS, C7LBDBS and C7LBDCS) were tested against three different target oligonucleotide sequences, which contained one C7 half site or two C7 half sites either in palindromic or direct repeat orientation.

5 Oligo 1: gat cca aag tcg cgt ggg cgc agc gcc cac gcg atc aaa ga (SEQ ID NO: 48)

10 Oligo2: gat cca aag tcc agg cga gcg cgt ggg cgg cag atc aaa ga (SEQ ID NO: 49)

15 Oligo3: gat cca aag tcg cgt ggg cgc agg cgc gag cgt ggg cgg atc aaa ga (SEQ ID NO: 50) Gel shift assay conditions were the same as the standard protocol described above. The results showed that C7LBDAS and C7LBDBS were able to bind to both oligonucleotides containing two C7 half sites, but not to the oligo containing only one half site. C7LBDCS bound weakly or not at all to all three targets.

20 Fusion proteins C7LBDA and C7LBDB bound to the probe containing a palindrome (two inverted half sites) as a single form and in equal amount to the C7VP control, while C7LBDC showed no detectable binding. In contrast, the fusion proteins C7LBDA and C7LBDB did not bind to the oligonucleotide containing only one C7 site, while C7VP bound only once, as expected. C7LBDA and C7LBDB bound equally to the oligo 1 and oligo 3, which contain two sites as inverted repeats with 3 intervening spaces or direct repeats with 9 intervening spaces,

25 respectively. These data indicate that the ZFP-LBD fusion proteins dimerize and bind preferentially to DNA containing two C7 half sites, but that the exact orientation and spacing of the half sites is not critical. This flexibility in DNA binding site orientation may reflect the lack of a dimerization function in the C2H2 domains, but it is noteworthy that wild type estrogen receptor has also been shown to bind a variety of response

elements differing from the consensus ERE, including inverted and direct repeats.

To further confirm the homodimer binding stoichiometry of the ZFP-LBD fusion proteins and to demonstrate their DNA sequence specificity, 5 the following experiment was conducted. A second ZFP-LBD fusion protein was constructed using the C2H2 zinc finger domain E2C(HS1), which binds to a recognition sequence 5'-GGG GCC GGA g 3' that differs in six out of nine base pairs from the C7 binding site. (Note that the lower case g denotes a 10th base that makes a minor contribution to 10 the protein:DNA contact affinity.) Maps of specific examples constructed in the expression vector pcDNA3.1 are shown in FIG. 11 (E2CLBDAS) (SEQ ID NO: 11) and FIG. 12 (E2CLBDBS) (SEQ ID NO: 12).

Oligonucleotides were prepared containing an inverted repeat of 15 two C7 sites, two E2C sites, or a mixed heterodimeric site of one C7 and one E2C half site. Two fusion proteins having different DNA binding domains (C7 or E2C) were tested for their DNA binding specificity against three oligonucleotides containing palindromic binding sites specific for C7, E2C or the combination of the two.

C7 oligo: gat cca aag tcg cgt ggg cgc agc gcc cac gcg atc aaa ga (SEQ 20 ID NO: 51)

C7/E2C oligo: gat cca aag tcg cgt ggg cgc act ccg gcc ccg atc aaa ga (SEQ ID NO: 52)

E2C oligo: gat cca aag tcg ggg ccg gag act ccg gcc ccg atc aaa ga (SEQ ID NO: 53)

25 Gel shift assays were performed according to the standard protocol described above.

The results showed that C7LBD fusion protein only binds strongly to the oligonucleotide containing two C7 sites, but not to either the 2 x E2C probe or the C7/E2C probe. Likewise, the E2C-LBD chimeric protein only binds strongly to the 2 x E2C probe. Finally, neither ZFP-LBD construct binds to the oligonucleotide with the heterodimeric site. When

the two proteins were mixed in equal amount, a C7LBD and E2CLBD heterodimer was formed. The heterodimer binds to the heterodimeric probe. These results confirm that the ZFP-LBD fusion proteins are binding DNA preferentially as dimers. Furthermore, these data demonstrate good DNA binding specificity between fusion proteins with different C2H2 binding site preferences.

EXAMPLE 5

Ligand-dependent Regulation of Transgene Expression by ZFP-LBD Fusion Proteins

In order to evaluate the ability of the fusion proteins C7LBD A, B, and C to regulate transgene expression, a standard co-transfection reporter assay was performed. A reporter construct, henceforth known as 6x2C7pGL3Luc, containing six copies of a directly repeated C7 binding site (6x2C7) inserted upstream of an SV40 promoter fragment and reporter gene encoding firefly luciferase (pGL3Pro; Promega) was transfected along with the designated fusion protein and assayed as described below.

Cultured cells (HeLa, Cos, Hep3B or other) were seeded at 5×10^4 cells/well in a 24 well plate prior to the day of transfection in DMEM Phenol-free media, supplemented with L-glutamine and 5% (v/v) charcoal-dextran stripped Fetal Bovine serum (sFBS). Cells were transfected using the Qiagen Superfect Transfection method. For each well 1 μ g of total DNA, containing 0.5 μ g luciferase reporter plasmid (6X2C7pGL3proluc), 0.1 μ g of chimeric activator DNA (e.g., C7LBDA, C7LBDB, or C7LBDC) unless otherwise indicated, and 0.4 μ g of an inert carrier plasmid DNA (p3Kpn), was mixed with 60 μ L of DMEM phenol-free/serum free media, and 5 μ L of Superfect reagent. In general, about 10 ng to about 0.5 μ g of chimeric activator DNA was used for each well.

The mixture was vortexed for 10 seconds and incubated at room temperature for 10 minutes, followed by the addition of 350 μ L of DMEM phenol-free 5% sFBS media. Cells were washed once with Dulbecco's

phosphate buffered saline (DPBS) and the transfection mixture placed on the cells. Cells were washed once with DPBS following a 2.5 hour incubation at 37 degrees Celsius, and re-fed with DMEM Phenol-free 5% sFBS media.

5 At approximately 24 hours post-transfection, cells were treated with an inducing agent, 17 β -Estradiol or 4 OH-Tamoxifen as indicated, each at 100nM final concentration in DMEM Phenol-free 5% sFBS. Cells were harvested 24 hours later by washing once with DPBS and adding 200 μ L 1X reporter lysis buffer (Promega). Plates were frozen at -80 ° C
10 and thawed at room temperature for 1.5 hours on an orbital shaker at 100 RPM. After allowing for cellular debris to settle, lysate was diluted 1:10 with 1X reporter lysis buffer, and 10 μ L transferred to 96 well opaque plates. Plates were analyzed with a Tropix TR717 Microplate Luminometer using firefly luciferase substrate (Promega).

15 The ability of C7LBD short form chimeric proteins A, B, and C to regulate reporter gene expression in an estrogen-dependent manner was studied in Cos and HeLa cells. The constitutive activators C7VP16 and 2C7VP16 were used as positive controls. The results show that the three ZFP-LBD fusion proteins gave a similar profile in Cos and Hela cells. All
20 three ZFP-LBD fusion proteins had an estrogen dependent effect on the luciferase reporter gene. The characteristic pattern is that A has greater total activity than B and B has greater total activity than C. Likewise, the basal or ligand-independent effect of these proteins on the reporter gene follows a similar pattern; A>B>C. The estrogen dependent effect on
25 gene expression ranged from two-fold to nine-fold in these experiments.

30 The regulation of luciferase reporter gene by the C7LBD long and short form fusion proteins was compared in Cos cells. The results indicate that the long form fusion proteins, which contain the estrogen receptor F region, have a higher basal and ligand-independent effect on the reporter gene than the short form. As a result, the long fusion proteins give lower fold induction. This result may be due to an

enhanced, but ligand-independent, transactivation activity in the F region that works synergistically with the heterologous VP minimal domain trimer. Alternatively, this result could be due to the difference in spacing, as a result of the intervening F region, between the VP activation domain and the estrogen receptor ligand binding domain of the recombinant proteins.

In order to evaluate the role of the composition of the heterologous transactivation domain on the activity of the C7LBD fusion proteins, the VP minimal domain trimer was replaced with either the carboxy terminal activation domain from human STAT-6 (amino acids 660 — 847) or the full length VP16 activation domain of approximately 77 amino acids (residues 413 — 490) (FIG. 13). In constructs with full length VP16, the transactivation domain was added either native, or in conjunction with an SV40 nuclear localization peptide sequence at the amino terminus of the VP16. C7LBD fusion proteins A or B containing different transactivation domains (TA2, STAT6C, VP16 and NLSVP16) were constructed and evaluated for their effects on gene activation and ligand induction. The construct, shown schematically and abbreviated above, includes the following:

- 20 1. C7ASTA2, C7BSTA2: C7LBD A or B short form with the VP16 minimal domain trimer.
2. C7BS-STAT: C7LBDB short form with the STAT6 carboxy activation domain.
3. C7BS VP16: C7LBDB short form with full length VP16 activation domain.
- 25 4. C7AS nlsVP16: C7LBDA short with full length VP16 preceded by a nls.

Assays were performed with HeLa cells transfected with 0.5 ug of 6x2C7pGL3Luc reporter and 0.1 μ g regulator, Luc activity was determined as previously described. When the human STAT6 transactivation domain was used to replace the TA2 VP minimal domain

trimer, the same low basal activity and 9 fold ligand dependent induction of transgene, two-fold less than with the TA2 domain, was obtained.

The incorporation of NLS upstream of the full length VP16 (FIG. 24, C7ASnlsVP16) greatly increased the folding induction compared to TA2 or VP16 without the NLS, but the total activity was significantly decreased. When the full length VP16 domain was used, it gave about 2 fold higher total activity, but high basal activity resulting in weaker ligand dependent induction (3-fold).

EXAMPLE 6

10 **Ligand-Independent Activity of C7-PBD-VP16 Constructs Depends On The Structure of the Reporter Constructs**

In initial tests, the C7-PBD-VP16 construct showed the high basal (i.e. ligand-independent) activity. Thus, C7-PBD-VP16 was compared to the original, Gal4-based construct GL914VPC', which reportedly had a very low basal activity. When the GL914VPC' protein was tested on a 6xGal4-SV40 promoter-luciferase reporter, it displayed even higher basal activity than C7-PBD-VP16. Variation of effector/reporter ratios had no effect on the basal activities in both systems. It was discovered, however, that the ratios for optimal induction were different for GL914VPC' and C7-PBD-VP16, namely 1/30 and 1/10, respectively.

Other possible sources of ligand-independent activity were examined. Commercially available fetal calf serum (FCS) batches are known to contain estrogen or estrogen-like activities. Since it was possible that the presence of progesterone-agonistic activities in the serum was the cause for the high basal activities, the FCS was "stripped" of steroids using dextran-coated charcoal. However, side-by-side comparison of stripped and non-stripped serum showed no detectable difference in the basal activity of the switch constructs. Lipid-based transfection reagents such as Lipofectamine™ can also have significant agonistic activity on steroid receptors. Thus, the non-lipid transfection

reagent Superfect™ from Qiagen was used as an alternative, and compared to Lipofectamine™.

No reduction of the basal activities was observed. For all the assays described above, HeLa cells were used. However, the use of HepG2 cells, 5 which were used in the original study with GL914VPc', brought no improvement.

The reporter p17x4TATA-luc, used in the original studies on Gal4, contains four Gal4 dimer binding sites upstream of a TATA box.

GL914VPc' had a very low basal activity on this reporter, and was 10 inducible by RU486. An equivalent reporter, pGL3TATA/10xC7, was therefore constructed to test C7-PBD-VP16. While the basal activity using a reporter construct having TATA reporter was still higher than in the Gal4 system, basal activity was clearly lower than using the SV40 promoter-containing pGL3prom/10xC7. Two additional reporters with 15 minimal CMV promoters, pGL3minCMV/6xGal4 and pGL3minCMV/10xC7, were also constructed. The basal activity of the corresponding switch proteins was as high on these reporters as on the SV40 promoter containing reporters.

These results indicate that GL914VPc' and C7-PBD-VP16 were 20 constitutively located in the nucleus and able to bind to their target sites, either as monomers or as dimers. However, unless bound to ligand the fusion proteins are only able to activate transcription in the context of more than a TATA box, i.e. a SV40 promoter or a minimal CMV promoter. If there is only a TATA box, ligand binding presumably 25 associated with a conformational change is required for efficient activation of transcription.

It was found that ligand-independent basal activity is also cell type specific. C7-PBD-VP16 had an even lower basal activity on the TATA reporter in NIH/3T3 cells than it had in HeLa cells.

30 Since C7-PBD-VP16 appears to be constitutively translocated to the nucleus, the SV40 nuclear localization signal (NLS) between PBD and

VP16 domains was removed in the hope of making nuclear translocation more ligand dependent. The resulting construct, C7-PBD-VP16noNLS, was then tested on the pGL3prom/10xC7 reporter. However, transcriptional activation was no more RU486-dependent than in the case of C7-PBD-VP16 as shown by an unchanged basal activity. The construct C7-PBD Δ NLS-VP16noNLS was made in which the small remaining part of a natural SV40-like NLS at the N-terminus of the PBD (aa 640-644) is also removed.

EXAMPLE 7

10 Optimizing Spacing and Orientation of the DNA Binding Domain Half-Sites

Naturally occurring steroid hormone receptors typically bind to an inverted repeat, or palindromic SRE. However, it has been shown in several cases that there is some flexibility in binding. Direct repeats and everted repeats can also serve as response elements. To determine the optimal spacing and orientation of the two half-sites for binding of a steroid receptor-based switch construct a total of eighteen C7 dimer TATA-luciferase reporter constructs were prepared. Six C7 dimers each in direct, inverted and everted repeat orientation, with spacers of 0 to 5 intervening bases. A test of the RU486-responsive C7-PBD-VP64 protein on each of these reporter constructs revealed that indeed there was quite some flexibility, since RU486 inducible activation was observed with each of the reporters (Tables 7-9, below; values listed are means of two determinations and the standard deviation). There were clear differences in the degree of responsiveness of each of the reporters.

25 A direct repeat of two C7 sites without any spacing displayed the most favorable properties. This is particularly important, indicating the ability to target (GNN)₆ sites using homodimeric and heterodimeric recombinant ligand-responsive transcription factors.

30 Further tests on the RU486-responsive VP64-C7-PR protein and the tamoxifen-responsive VP64-C7-ER protein, on each of these reporter constructs also revealed some flexibility, since ligand-inducible activation

was observed with each of the proteins on each of the reporters. However, the most favorable properties were observed with the VP64-C7-ER protein on the direct and everted repeats with a spacing of 3bp. Direct repeat with a spacing of 5bp was also more or less reasonable, permitting targeting of the *erbB-2* promoter with a 3 finger construct (see below).

Further studies have shown that binding of a C7/Cf2-PBD-VP64 heterodimer to a C7-Cf2 TATA reporter, with one binding site each for C7 and Cf2 without spacing, provides about a two-fold ligand-dependent change in transcription.

Table 7

	C7-PBD-VP64	Direct Repeats	
		Mean	STD DEV
15	C7c7	4081	511
	C7c7 + RU486	20018	2090
20	C7ac7	3383	396
	C7ac7 + RU486	8205	2064
	C72ac7	3417	348
	C72ac7 + RU486	8169	634
25	C73ac7	3269	1550
	C73ac7 + RU486	5138	2319
	C74ac7	3966	298
	C74ac7 + RU486	6945	1377
	C75ac7	2597	416
	C75ac7 + RU486	5460	207

TABLE 8		
	C7-PBD-VP64	Inverted Repeats
	Mean	STD DEV
5	C77c	2921 1368
	C77c + RU486	10811 1596
	C7a7c	4342 153
	C7a7c + RU486	9534 2943
	C72a7c	6964 573
10	C72ac7 + RU486	19186 3284
	C73a7c	7132 5208
	C73a7c + RU486	12844 171
	C74a7c	3502 416
	C74a7c + RU486	8855 2379
15	C75a7c	4704 105
	C75a7c + RU486	12444 2117

Table 9		
	C7-PBD-VP64	Everted Repeats
	Mean	STD Dev
20	7cc7	8750 1839
	7cc7 + RU486	17377 1335
	7cac7	6029 613
25	7cac7 + RU486	13599 2014
	7c2ac7	7880 1720
	7c2ac7 + RU486	20825 8197
	7c3ac7	9670 1187
	7c3ac7 + RU486	21491 274
30	7c4ac7	6974 441
	7c4ac7 + RU486	8896 2455
	7c5ac7	6892 388
	7c5ac7 + RU486	13124 3490

EXAMPLE 8

C7-PBD-Repressor Domain Fusion Constructs.

To evaluate the use of PBD fusion proteins as regulatable transcriptional repressors, C7-PBD was fused to a number of repressor domains (Table 10, below). When tested in luciferase reporter assays, many repressor constructs had no significant activity. C7-PBD-KK (containing a dimer of two KRAB-A boxes) reproducibly led to a 25-50% repression, which was largely RU486-dependent. A much stronger repression which, however, was largely RU486-independent was observed with a C7-PBD-SKD construct.

EXAMPLE 9

Regulation Of *erbB-2* Promoter Activity With Three Finger-PBD-VP64 Homo-/Hetero-Dimers

The C7-PBD-VP16 switch protein was able to regulate 10xC7 reporter constructs, which contain 10 direct repeats of C7 sites with a spacing of 5bp (see above), indicating that a switch dimer can bind to direct repeats with this specific spacing. To evaluate the potential use of homo- and hetero-dimeric three finger-PBD fusion proteins for the ligand-dependent regulation of *erbB-2* promoter activity, the promoter region was screened for the presence of (GNN)₃N₅(GNN)₃ motifs. Four dimer target sites (E2E, E2F, E2G, and E2H) were identified. E2E overlaps with the 18bp E2C target sequence and could serve as a binding site for a homodimer. The other three sites have the potential to serve as heterodimer binding sites. The seven required three finger proteins were generated by F2 stitchery and analyzed for binding by ELISA (Table 11, below). *erbB-2*-specific switch constructs were then generated by fusion of each three finger protein to PBD-VP64, and tested for their ability to regulate *erbB-2* promoter activity. The values are mean and standard deviation of duplicate measurements. Only the heterodimeric E2F-PBD-VP64 switch led to a detectable regulation of the *erbB-2* promoter. This

regulation was not RU486-dependent, consistent with the high basal activities of C7-PBD-VP16 and C7-PBD--VP64 proteins.

Table 10
Progesterone Receptor Based Ligand-Responsive Transcription Factors

	DNA Binding Domain	Ligand Binding Domain	Transcription Effector Domain
5	C7	hPR (aa 640-914)	VP16
	C7	hPR (aa 640-914)	VP64
	C7	hPR (aa 640-914)	KRABa
	C7	hPR (aa 640-914)	Mad
10	C7	hPR (aa 640-914)	Mad-Mad
	C7	hPR (aa 640-914)	KRABa-Mad
	C7	hPR (aa 640-914)	Mad-KRABa
	C7	hPR (aa 640-914)	Deactylase
	C7	hPR (aa 640-914)	SKD
15	2C7	hPR (aa 640-914)	VP16
	2C7	hPR (aa 640-914)	VP64
	E2E 3F	hPR (aa 640-914)	VP64
	E2F 3F	hPR (aa 640-914)	VP64
	E2G 3F	hPR (aa 640-914)	VP64
20	E2H 3F	hPR (aa 640-914)	VP64
	E2C(SP1) 6F	hPR (aa 640-914)	VP16
	E2C(SP1) 6F	hPR (aa 640-914)	VP64
	E2C(SP1) 6F	hPR (aa 640-914)	KRABa
	E2C(SP1) 6F	hPR (aa 640-914)	Mad
25	E2C(SP1) 6F	hPR (aa 640-914)	KRABa-KRABa
	E2C(SP1) 6F	hPR (aa 640-914)	Mad-Mad
	E2C(SP1) 6F	hPR (aa 640-914)	KRABa-Mad
	E2C(SP1) 6F	hPR (aa 640-914)	Mad-KRABa

Table 11

Target	Target Sequence	Binding	Mean Basal Activity	STD DEV Basal Activity	Mean RU486 Activity	STD DEV RU486 Activity
	Control pcDNA 3.1		17209	1878		
E2C-HS1	ggg-gcc-gga	good				
E2C-HS2	gcc-gca-gtg	good				
E2E	gcc-gga-ggc	none	18259	140	15893	2083
E2F-HS1	gag-gag-ggc	good	61401	25291	54986	19240
E2F-HS2	gag-gaa-gta	?				
E2G-HS1	ggg-gcc-ggg	weak	25982	5444	12394	139
E2G-HS2	ggc-gca-gta	weak				
E2H-HS1	ggc-gcg-ggg	weak	15374	844	15374	537
E2H-HS2	ggt-gct-gcg	none				

EXAMPLE 10

15 **Estrogen And Progesterone Receptor Fusion Proteins With N-Terminal Effector Domains**

Recombinant ligand-responsive polypeptides were constructed using an estrogen receptor (ER) ligand binding domain (EBD). A Myc-ER fusion construct was obtained from Eliane Muller and used as a source of the EBD coding region. Rather than containing the human wild type amino acid sequence, Myc-ER contains a point mutation (aa 282-599, G525R) mouse EBD which has been shown to no longer bind estrogen, but bind the estrogen antagonist 4-OH tamoxifen, and paradoxically becomes activated by it. This has advantages for *in vivo* applications and for tissue culture experiments, not only because serum contains estrogen but also because phenol red present in all tissue culture media acts as an estrogen agonist.

The VP16-C7-ER, VP16-NLS-C7-ER, and VP16-C7-NLS-ER fusion constructs were prepared as described above. In parallel, an analogous set of progesterone receptor (PR) variants was also prepared (VP16-C7-PR, VP16-NLS-C7-PR, and VP16-C7-NLS-PR. The PBD in these

constructs encompasses aa 640-914 and therefore lacks the partial natural NLS (aa 640-644).

Each of these constructs was tested in a luciferase assay and compared to C7-PBD-VP16, using pGL3prom/10xC7 as a reporter. Not only did all these PR constructs have a higher activity in the presence of RU486 than C7-PBD-VP16, but the completely NLS-free VP16-C7-PR also had a significantly lower basal activity. This resulted in a dramatically improved ligand-dependent induction, 26-fold vs. 6-fold in this particular experiment. Tamoxifen-induced activity of the ER constructs was roughly four times higher than RU486-induced activity of the PR variants. Ligand-dependent induction was better; 43 fold for VP16-C7-ER.

The VP16 domain in VP16-C7-PR and VP16-C7-ER has been replaced by the following effector domains: the activator VP64, and the repressors KK (KRAB-A box dimer), MM (dimer of the Mad sin3 interaction domain) and SKD. The VP64 variants are useful, for example, in studies to determine the optimal spacing and orientation of the two half-sites, using the above-mentioned C7 dimer-TATA luciferase reporters

EXAMPLE 11

Targeting natural promoters using 3 Finger proteins fused to nuclear hormone LBDs

The following target sequences for 3 Finger switch homo- and hetero-dimers have been identified in the human *erbB-2* (E2) and integrin β_3 (B3) promoters:

E2E	<u>GCC GGA GCC</u> ATGGG GCC GGA GCC	direct repeat,
25	5bp spacing homodimer (SEQ ID NO: 54)	
B3D	CGC TCC CTC TCA GGC GCA GGG	everted repeat,
	3bp spacing, heterodimer (SEQ ID NO: 55)	

B3E GGC GCC CAC TGT GGG GCG GGC everted repeat,
5 3bp spacing, heterodimer (SEQ ID NO: 56).

EXAMPLE 12

Targeting Natural Promoters Using Six Finger Proteins Fused To Nuclear Hormone Ligand Binding Domains

The "6 Finger heterodimer"

Regulation of a 6 finger protein binding to a single 18bp site using any of the formats described have been unsuccessful. Similarly, a C7-PBD-VP64 protein did not activate a TATA reporter containing only a single C7 site. As an alternative, heterodimer constructs were prepared in which only one of the dimerization partners contains a DNA binding domain, while the other contains an effector domain.

The formats were as follows:

(1) E2C-PR // PR-VP64

(2) E2C-ER // ER-VP64

All four fusion constructs were fully sequenced and tested in a luciferase assay for their ability to regulate the *erbB-2* promoter in a ligand-dependent manner. It was found that the PR 6 Finger heterodimer was inactive; a similar observation was made with an C7-RxR // EcR-VP16 heterodimer. In contrast, the E2C-ER // ER-VP64 heterodimer had some activity, and the addition of Tamoxifen lead to a roughly three-fold upregulation of promoter activity. Variations in the ratio of the two heterodimerization partners led to an increased inducability, up to total of 5.3-fold.

The coding region for RXR (mammalian) and EcR (*Drosophila*) were PCR amplified from pVgRXR (Invitrogen) using the primers listed below and AmpliTaq DNA Polymerase (Hoffmann-LaRoche). Forward and backward primers were chosen to allow construction of the constructs. The cycling conditions were 2'/94° C C; 25 x (30''/94° C — 30''/60° C — 2'/72° C); 10'/72° C. The PCR product was purified with the Quiagen PCR prep kit, cut with the indicated restriction endonucleases and ligated

into a modified eukaryotic expression vector pcDNA3 (Invitrogen; see, also, Beerli *et al.* (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:14628-14633) to yield the constructs in FIG. 14.

Primers:

5 (FseI)-RXR: (SEQ ID NO: 57)
GAGGAGGAGGGCCGGCCGGGAAGCCGTGCAGGAGGAGCGGC
RXR-(Ascl): (SEQ ID NO: 58)
GAGGAGGAGGGCGCGCCCAGTCATTGGTGC GGCGCCTCCAGC
RXR-(PacI): (SEQ ID NO: 59)
10 GAGGAGGAGTTAATTAAAGTCATTGGTGC GGCGCCTCCAGC
(FseI)-EcR: (SEQ ID NO: 60)
GAGGAGGAGGGCCGGCCGGGTGGCGGCCAAGACTTGTAAAGAAGG
(Sfil)-EcR: (SEQ ID NO: 61)
GAGGAGGAGGGCCCAGGC GGCGGTGGCGGCCAAGACTTGTAAAGAA
15 GG
EcR-(Ascl): (SEQ ID NO: 62)
GAGGAGGAGGGCGCGCCGGCATAACGTCCCAGATCTCCTCGAG

Exchange of zinc finger and effector domains

20 After digestion with the restriction endonuclease Sfil the C7 3-finger protein was replaced with the 6-finger proteins E2C, B3B, B3C2 and 2C7 by standard cloning procedures. After digestion with the restriction endonucleases Ascl and PacI the activation domain VP16 was replaced with the activation domain VP64 and the repression domains KK and SKD.

25 *Luciferase assays*

For all transfections, HeLa cells were plated in 24-well dishes and used at a confluence of 40-60%. Typically, 175 ng reporter plasmid (pGL3-promotor constructs or, as negative control, pGL3basic) and 25 ng effector plasmid (zinc finger constructs in pcDNA3 or, as negative control, empty pcDNA3.1) were transfected using the Lipofectamine reagent (Gibco BRL). Cell extracts were prepared approximately 48 hours

after transfection. Luciferase activity was measured with the Promega luciferase assay reagent in a MicroLumat LB96P luminometer (EG&G Berthold).

Bombyx mori EcR

5 A plasmid (LNCVBE) containing the coding region for *Bombyx mori* EcR was obtained from F. Gage. *Bombyx mori* EcR is PCR amplified from this plasmid using the primers listed below and AmpliTaq DNA Polymerase (Hoffmann-LaRoche). Forward and backward primers were chosen to allow construction of the constructs corresponding to FIG. 14
10 but replacing *Drosophila* EcR by *Bombyx mori* EcR.

(FseI)-BE: (SEQ ID NO: 63)

GAGGAGGAGGGCCGGCCGGAGGCCTGAATGTGTCATACAGGAGCCC

(Sfil)-BE: (SEQ ID NO: 64)

GAGGAGGAGGGCCCAGGCAGGCCAGGCCTGAATGTGTCATACAGGAGCCC

15 BE-(Ascl): (SEQ ID NO: 65)

GAGGAGGAGGGCGCGCCCTCCGCCACGTCCCAGATCTCCTCGAG

C7-R-VP16 // C7-E-VP16

This heterodimer was examined on two reporters, one containing 10 C7 sites and one containing 6 2C7 sites, and in two cell lines, HeLa
20 and NIH. In all cases the C7-R-VP16 construct alone showed a high activation of transcription (840-fold) that did not depend on the presence of Ponasterone A. However the C7-E-VP16 construct showed a very little activation of transcription on its own. C7-R-VP16 // C7-E-VP16 together showed the same behavior as C7-R-VP16 alone.

25 C7-R // E-VP16

In this heterodimer, the activation domain on RXR is dropped to eliminate the basal activation observed above. EcR has no DNA-binding domain to render activation dependent on the presence of DNA-bound RXR. This heterodimer was tested with the 3-finger protein C7 on the
30 10C7 reporter and with the 6-finger protein E2C on the E2P reporter that

contains a single E2P binding site. In both cases no significant activation could be observed.

C7-R // C7-E-VP16

To combine the low basal activity of C7-R // E-VP16 with the high activation seen with C7-R-VP16 // C7-E-VP16, the activation domain on RXR was dropped but the zinc finger protein on EcR was retained. In this set-up, on a 6x2C7 reporter, a 5-fold activation with very low basal activity was observed. Similar constructs using the more powerful VP64 activation domain have also been made.

E2C- ER // ER-VP64

This heterodimeric construct showed 5.3 fold tamoxifen-dependent activation at ratios of 6.7/60 and 2.2/60 of the erbB-2 promoter.

E2C- ER // ER-KRAB

This heterodimeric construct showed 2.9 fold tamoxifen-dependent repression of the erbB-2 promoter at a ratio of 1/10.

B3B/B3C2-ER // ER-VP64

This six finger heterodimeric construct showed 4.5 - 7.8 fold tamoxifen-dependent activation of the β 3 promoter.

EXAMPLE 13

Regulation Of Endogenous ErbB-2 Gene Expression Using Adenovirus-Mediated Delivery Of E2C-KRAB

Adenovirus vectors can be produced at very high titers, which makes them useful for gene therapy applications. To demonstrate the use of the E2C-KRAB repressor protein in animal models, E2C-KRAB (and, as a control, 2C7-KRAB) encoding adenoviruses were generated. The method for adenovirus production is described in detail, for example, in He *et al.* (1998) *Proc. Natl. Acad. Sci. U.S.A.* 95:2509-2514.

Briefly, the zinc finger coding regions were excised from the pMX/E2C-KRAB and pMX/2C7-KRAB bicistronic retrovirus plasmids by BamH1-Not1 digest. The resulting fragments were then subcloned into the Bgl2-Not1 sites of pAdTrack-CMV. After linearization with Pme1,

pAdTrack plasmids were co-electroporated with circular pAdEasy-1 into BJ5183 cells. This bacterial strain is not *recA* and therefore allows homologous recombination between the 2 plasmids. Electroporated cells were then plated onto Kan plates. Only plasmids that have recombined together provide Kanamycin resistance, because this marker is only present on pAdTrack. After screening to distinguish recombinants from background (due to incomplete linearization of pAdTrack plasmids), the linear adenovirus vector genomes were released from the recombinant pAdEasy/E2C-KRAB and pAdEasy/2C7-KRAB plasmids by Pac1 digest.

The linearized vectors were then transfected into 293 cells. This cell line makes the Adeno E1A and E1B proteins, which have been deleted from the vector and are required for replication.

EXAMPLE 14

Modifications To The Estrogen Receptor Ligand Binding Domain Improve Ligand Dependent Induction And Ligand Selectivity.

Single amino acid mutations in the estrogen receptor ligand binding domain can have a significant effect on the basal and ligand dependent level of gene activation. For example, a glycine to valine substitution at estrogen receptor residue 400, has been described as a destabilizing or temperature sensitive mutation (White (1997) *Adv. Pharmacol.* 40:339-367; Aumais *et al.* (1996) *J. Biol. Chem.* 272:12229-12235). The effect of this mutation on the properties of the fusion proteins was tested. The general methods for constructing fusion proteins with altered amino acids is described below.

Mutagenesis of the fusion proteins C7LBDA and C7LBD**b** was performed using oligonucleotide mediated site directed mutagenesis (Stratagene; Quikchange Site-Directed Mutagenesis Kit) to either substitute arginine for glycine at amino acid 521 (G521R-human estrogen receptor nomenclature) or a valine residue for glycine at amino acid 400 (G400V). The sequences of the oligonucleotides used for G521R mutagenesis were

GTACAGATGCTCCATGCGTTACTCATGTGCC (SEQ ID NO: 66) for
the noncoding strand and
GGCACATGAGTAACAAACGCATGGAGCATCTGTAC (SEQ ID NO: 67) for
the coding strand, where the nucleotide in bold represents the change
from the wild type sequence.

5 The sequences of the oligonucleotides used for G400V
mutagenesis were CCATGGAGCACCCAGTGAAGCTACTGTTGC (SEQ ID
NO: 68) for the coding strand, and,
GCAAACAGTAGCTTCACTGGGTGCTCCATGG (SEQ ID NO: 69) for the
10 noncoding strand, where the nucleotide in bold represents the change in
sequence from wild type.

Templates were added at 10 ng to 50 ng per reaction with 125 ng
of each primer in 10mM KCl, 10mM (NH₄)₂SO₄, 20mM Tris-HCl (pH 8.8),
2mM MgSO₄, 0.1% Triton X-100, 0.1 mg/ml BSA, dNTP mix, and 2.5U
15 PfuTurbo™ DNA polymerase. The reactions were carried out on a Perkin
Elmer GeneAmp PCR system 9600 thermal-cycle using an initial
temperature of 94 degrees Celsius for 30 seconds to denature the
template, followed by 12 cycles at 95 degrees Celsius for 30 seconds, 55
degrees Celsius for 1 minute, and 68 degrees Celsius for 4 minutes, with
20 a single round of extension at 72 degrees Celsius for 2.5 minutes. PCR
samples were treated with 10U *DpnI* for 1hr at 37 degrees Celsius to
digest the non-mutagenized parent template.

DH5 α supercompetent Epicurean Coli® XL-1 cells were transformed
by combining 1 μ L of the *DpnI* treated PCR samples with 50 μ L of the
25 cells in chilled Falcon 2059 tubes, incubated on ice for 30 minutes, heat
shocked at 42 degrees Celsius for 45 seconds and chilled on ice for 2
minutes. A 500 μ L aliquot of SOC media pre-warmed to 42 degrees
Celsius was added to the transformation reaction and incubated for 1
hour at 37 degrees Celsius with shaking. The transformed cells were
30 plated onto LB plates containing 100 μ g/ml ampicillin and incubated for at
least 16 hours.

Mutation efficiency was determined by altering a nonsense codon in a β -galactosidase expression plasmid to glutamine and determining expression of β -galactosidase, as evidenced by IPTG/X-Gal plates. Approximately three clones for each mutation were selected for restriction enzyme digestion to check for template integrity, followed by dideoxynucleotide sequencing of the entire coding frame to confirm the desired mutation.

C7LBD (short) chimeric regulators A, B, and C with and without the G400V mutation in the estrogen receptor LBD were compared for their ability to induce expression of the 6x2C7pGL3Luc. As observed previously, the total activity of the three fusion proteins has the relationship A>B>C; this relationship was maintained with and without the G400V mutation. The pattern of basal expression was dramatically altered by the G400V mutation. The basal or ligand independent effect of the three C7LBD regulators with the G400V mutation is reduced to nearly the level of the reporter plasmid alone. As a result, the fold ligand dependent induction dramatically increases, for example from 10 fold to 420 fold for C7LBDA.

It has previously been observed with fusion proteins containing an estrogen receptor ligand binding domain, that activity could be induced by use of not only the natural agonist estrogen (E2) but also synthetic anti-estrogens such as 4-OH tamoxifen (Littlewood *et al.* (1995) *Nucl. Acids Res.* 23:1686-1690; Danielian *et al.* (1993) *Mol. Endocrinol.* 7:234-240). The ability of the C7LBD fusion to be induced by 4-OH-tamoxifen was demonstrated.

The results of the study showed the ligand-dependent regulation of a luciferase reporter gene construct in HeLa cells using three recombinant molecular constructs, C7LBDAS, C7LBDBS, C7LBDCS with and without a G400V mutation in response to estrogen (E2) and 4 hydroxytamoxifen (OHT). In particular, the results showed that fusion proteins C7LBD B and C are induced equally well by 100 nM tamoxifen or estrogen. For

C7LBDA, tamoxifen appears to be approximately two-fold more active than estrogen itself.

Another mutation of interest in estrogen receptor LBD is a glycine to arginine substitution at amino acid 521 of human estrogen receptor. This mutation has also been described in the mouse estrogen receptor homolog at the equivalent site of residue 525. This mutation ablates responsiveness of the mutated LBD to estrogen, but still allows the binding of the anti-estrogen tamoxifen (Littlewood *et al.* (1995) *Nucl. Acids Res.* 23: 1686-1690; Danielian *et al.* (1993) *Mol. Endocrinol.* 7:234-240). The effect of the G521R mutation on the activity of the C7LBD regulators was tested. C7LBDB was compared to C7LBDB (G400V) and C7LBDB (G521R).

The results of the study showed the ligand-dependent regulation of a luciferase reporter gene construct in HeLa cells using three recombinant molecular constructs: C7LBDBS, C7LBDBS with a G521R mutation and C7LBDBS with a G400V mutation in response to estrogen (E2) and 4 hydroxytamoxifen (OHT). Similar to the effect observed with the G400V mutation, G521R significantly reduces the basal activity of the fusion protein regulator. But most importantly, now the C7LBDB(G521R) regulator is fully activated by 100 nM 4-OH-tamoxifen, but completely inactive in response to estrogen. Note that the G400V mutant is still fully activated by estrogen and tamoxifen.

To further investigate the effect of the G521R mutation, a series of different estrogenic compounds were evaluated for their ability to induce the C7LBD regulators. The activity of 100 nM for four compounds: estrogen (E2) and diethyl-stilbesterol (DES) are estrogenic agonists, 4-OH-tamoxifen and raloxifene (Ral) are non-steroidal anti-estrogens, or so-called SERMS (selective estrogen receptor modulators) were compared.

The study tested ligand-dependent regulation of a luciferase reporter gene construct in Hep3BL liver cells using recombinant molecular constructs C7LBDBS with a G521R mutation and C7LBDBS with a G400V

mutation in response to estrogen (E2) diethylstilbestrol (DES), 4-hydroxytamoxifen (4-OHT) and raloxifen (Ralox). The results showed that the G521R mutation selectively eliminates response to the agonists, but the non-steroidal synthetic ligands tamoxifen and raloxifen are still fully active.

EXAMPLE 15

Effect Of The Minimal Promoter Composition On Regulation Of Transgenes By ZFP-LBD Fusion Proteins

The composition of the minimal promoter used in reporter assays can dramatically effect the level of gene expression. Likewise, the activity of natural steroid receptors varies on different gene targets depending on the composition of their promoters. Reporter constructs containing 6x2C7 binding sites upstream of a minimal TATA box promoter fragment derived from the c-fos gene, referred to here as TATA were constructed to show the effect on the level of regulation. C7LBD A and B fusions without or with the G400V or G521R mutations were compared. As observed previously on the pGL3 SV40 promoter, the G400V and G521R mutations significantly decrease the basal activity of the chimeras compared to those without these mutations. Further, the G521R mutant is selectively activated by tamoxifen. On this weaker minimal promoter, estrogen is only a weak inducer, while 4-OH-tamoxifen is significantly better. This effect is even more pronounced on C7LBD A compared to B; on chimera C7LBDA (G400V), tamoxifen is at least 10 fold more active than estrogen.

An experiment was done to directly compare the relative activity of the C7LBD chimeras on reporter constructs containing the stronger pGL3 SV40 promoter or the weaker c-fos TATA box promoter.

The results of the study show that the ligand-dependent regulation of a luciferase reporter gene construct expressed from a minimal TATA promoter in Hep3BL liver cells using recombinant molecular constructs C7LBDAS and C7LBDBS with a G400V mutation and C7LBDBS with a

G521R mutation in response to estrogen (E2) and 4-hydroxytamoxifen (OHT). Three important observations can be made: 1) the absolute level of induced activity is about 10 fold higher on the SV40 than the TATA promoter 2) the basal activity of the fusions is also about 10 fold higher on the SV40 than on the TATA promoter, 3) while both promoters show strong fold induction by tamoxifen (492 X on SV40 and 132 X on TATA), estrogen is only a strong inducer of the SV40 but not the TATA promoter (177X vs 14X). These results indicate that a gene regulation system using these fusion proteins can be "tuned" by choice of an appropriate minimal promoter.

Target Selectivity of Different C2H2 DNA binding domains

Reporter constructs with 3 copies of direct repeats of the C7 binding site (GCG TGG GCG) or E2C binding site (GGG GCC GGA g) inserted upstream of the promoter region in pGL3Luc were used to evaluate target specificity two different ZFP-LBDBs fusion protein regulators. ZFP-LBDB short fusions were constructed containing either the C7 DNA binding domain or the E2C DNA binding domain and tested on the two different reporter constructs. The study was designed to show the effect of three direct repeats of either C7 or E2C binding sites inserted upstream of the promoter of a luciferase reporter gene construct in HeLa cells on estrogen-dependent gene expression using recombinant molecular constructs C7LBDBS and E2CLBDBS. Estrogen-dependent induction only occurs when the chimera's DNA binding domain (DBD) matches the binding sites in the reporter. The E2CLBD chimera shows no increase of luciferase activity on the 3x2C7 Luc reporter and visa versa for C7LBD on the 3xE2C reporter.

It was previously determined from DNA binding studies that the fusion protein regulators have an absolute dependence on the presence of two half sites within a "response element" in order to bind DNA. In order to determine the optimal orientation and spacing of the binding sites for gene activation, a series of different reporter constructs were assembled.

In order to determine the optimal target DNA spacing and orientation of the C2H2 binding sites for transgene induction, C7LBDBS was transfected into HeLa cells and assayed for basal and tamoxifen induced activity on a series of reporter constructs.

5 A series of different reporter constructs assembled in order to determine the optimal target DNA spacing and orientation of the C2H2 binding sites for transgene induction, C7LBDBS was transfected into HeLa cells and assayed for basal and tamoxifen induced activity on a series of reporter constructs diagramed above. Reporter constructs were
10 constructed by cloning double stranded oligonucleotides containing the various binding sites into the multiple cloning site of the pGL3Luc reporter. "Response elements" composed of direct, inverted (palindromic), and everted repeats of two C7 binding sites were compared; each response element was separated by two (2) bp except in
15 the control 6 X 2C7, where spacing was 5 bp. Several arrays of directly repeated single C7 sites were tested with various spacing. The data show that direct repeats and everted repeats are preferred over palindromic binding sites. Further, 6 C7 sites, each separated by 2 bp is comparable to the control element of 6 x 2C7, even though it contains
20 only half the number of individual C7 binding sites.

EXAMPLE 16

Construction and Evaluation of ZFP-LBD Fusion Protein Regulators Containing Arrays of Six C2H2 Zinc Fingers

Studies were performed to determine if DNA binding domains comprised of zinc finger arrays binding up to 18 bp of DNA could be substituted for the normal estrogen receptor DBD. The previous constructs, containing three finger arrays that bind nine bp are a fairly conservative replacement of the wild type estrogen receptor ligand binding domain that binds six bp for each receptor monomer. The possibility exists that if large DNA binding domains are fused to an LBD fragment, that these domains may prevent dimerization via the LBD
25
30

dimerization domain due to steric interference. However, since the six finger arrays already provide high DNA specificity and affinity, dimerization may be unnecessary for the DNA binding and activity of these fusions proteins. Fusion protein regulators were prepared by fusing the 2C7 six finger array to the three LBD fragments A, B, and C described above. FIG. 15 provides a schematic and description of the cloning step required to assemble 2C7LBDshort A, B, and C.

Protein binding to DNA was analyzed by gel shift assay. The electrophoretic studies used 2C7 recombinant molecular constructs using native PAGE and SDS PAGE analysis of binding to a DNA probe containing six 2C7 binding sites. In this experiment, the 2C7VP16 protein was used as a control and the P32-labeled DNA probe was the 6x2C7 fragment excised from the 6X2C7pGL3Luc. Sufficient 2C7VP protein was added to yield three distinct gel shifted products. When a similar level of protein for the 2C7LBD A, B, and C were applied, only a single weak band was observed. By comparison to the one and two copies bound bands for the 2C7VP16 control, the 2C7LBD band position suggests it is binding as a monomer. Furthermore, the weak level of binding compared to the 2C7VP16 control suggests the DNA binding affinity of the 2C7 domain is significantly reduced in the context of the LBD fusion protein. Results of *in vitro* expressed proteins by SDS-PAGE, indicated equal amounts of proteins expressed and the expected relative increase in size for the LBD A, B, and C forms.

The ability of the 2C7LBD A, B, and C fusion protein chimeric regulators to activate expression of the 6X2C7Luc reporter gene were evaluated essentially as described previously for the C7LBD studies. The results of the study show the ligand-dependent regulation of a 2C7 SV40 luciferase reporter gene construct in Cos cells using three recombinant molecular constructs, 2C7LBDAS (SEQ ID NO: 1), 2C7LBDBS (SEQ ID NO: 2), 2C7LBDCS (SEQ ID NO: 3), and a positive control, 2C7-Vp16. The results are similar to the data evaluating C7LBD in Cos cells. The

2C7LBD regulators give about two fold estrogen dependent induction over basal, with 2C7LBDA > B > C for both the total activation activity and the increased basal activity relative to reporter plasmid alone. Maps of the additional constructs are depicted in FIG. 16 - FIG. 22.

5

EXAMPLE 17

Construction and Evaluation of Additional Reporter Transgene Constructs

An inducible promoter was constructed based on binding sites for the 3 Finger protein N1. The promoter contains 5 direct repeats of N1 sites spaced by 3bp; the spacing between the 5 repeats is 6bp. (FIG.

10

23A)

Luciferase assay. HeLa cells were cotransfected with plasmids encoding the indicated fusion proteins and the N1 reporter construct. At 24h later, the cells were treated with 10nM RU486 (FIG. 23B) or 100nM Tamoxifen (FIG. 23C), respectively. At 48h post transfection, cell extracts were assayed for luciferase activity.

15

Another inducible promoter based on binding sites for the 3 Finger protein B3. The promoter contains 5 direct repeats of B3 sites spaced by 3bp; the spacing between the 5 repeats is 6bp (FIG. 24A).

20

Luciferase assay. HeLa cells were cotransfected with plasmids encoding the indicated fusion proteins and the B3 reporter construct. At 24h later, the cells were treated with 10nM RU486 (FIG. 24B), or 100nM Tamoxifen (FIG. 24C), respectively. At 48 h post transfection, cell extracts were assayed for luciferase activity.

EXAMPLE 18

25

Heterodimer Formation in Presence of Ligand

30

FIG. 25 shows the results of a luciferase assay showing RU486-induced formation of functional VP64-C7-PR/VP64-CF2-PR heterodimers. HeLa cells were cotransfected with the corresponding effector plasmids and TATA reporter plasmids (C7/CF2-dr0, C7 site 5' to a CF2 site, direct "repeat", no spacing; C7/C7-dr0, 2 C7 sites, direct repeat, no spacing).

At 24h later, the cells were treated with 10nM RU486. At 48h post transfection, cell extracts were assayed for luciferase activity.

EXAMPLE 19

5 **Construction and Evaluation of the Cys₂-His₂ Zinc finger DBD – ER LBD regulators in Adenoviral Vectors**

In order to efficiently deliver the two components of the regulatory system to mammalian cells, either *ex vivo* or *in vivo*, a series of adenoviral vectors were constructed. These vectors contained either the ZFP-LBD fusion protein regulator linked to the immediate early CMV promoter or the regulatable transgene, linked to the 6 x 2C7 array of C7 binding sites and the minimal promoter from SV40 or c-fos TATA as described previously. The fusion protein regulator vector and regulatable transgene vector are then be mixed at various ratios and delivered to cells or animals by standard methods.

15 Construction of an adenovirus vector is routine and generally, the procedure involves three main steps: first a shuttle plasmid containing the viral left ITR, viral packaging signal, a promoter element, a transgene of interest linked to the promoter element and followed by a poly adenylation sequence, and some additional DNA sequences, viral or non-viral, required for recombination is prepared. Second, this left end shuttle plasmid, along with the remainder of the viral genome (i.e. the right end of the vector) are transfected into a host cell and joined through DNA recombination to form a complete vector genome. This recombination step may result from sequence homology between the two vector halves or may be aided by the use of site specific recombinases such as Cre and their corresponding LoxP recombination sequences. Finally, the newly formed virus is amplified up and purified in a series of steps. The details of the construction of these vectors are briefly described below.

20

25

30 **Left end shuttle plasmid construction for ZFP-LBD Fusion Protein Regulators**

Shuttle plasmids containing the left viral ITR, CMV immediate ealy promoter and ZFP-LBD regulator were prepared in the plasmid pAvCVlx

(Figure 26). Note that this vector contains a loxP recombination site just downstream of the poly adenylation sequence. DNA encoding the intact reading frame for the chimeric regulators C7LBD As(G521R), C7LBD Bs(G521R), and C7LBD Bs(G400V) were excised from the appropriate pCDNA constructions, (see figures 4 and 5 for LBD As and LBD Bs constructs respectively) by digestion with restriction enzymes EcoRI and Not I. The ZFP-LBD DNA fragments were modified with Klenow to fill in the restriction site overhangs and blunt end ligated into the EcoRV at bp 1393 site of pAvCvIx to generate pAvCv-C7LBD As(G521R), pAvCv-C7LBD Bs(G521R), and pAvCv-C7LBD Bs(G400V).

10 Construction of Left end shuttle plasmids containing regulatable transgene cassettes

Two regulatable transgene cassettes were prepared. One contained the 6x2C7 binding sites and SV40 minimal promoter fragment linked to the Luciferase transgene as in pGL3 6x2C7-Luc (described in example 5). The second vector contained the 6x2C7 binding sites and c-fos TATA minimal promoter linked to a cDNA encoding murine endostatin fused to an amino terminal secretion signal. The complete sequence of this fusion protein is listed in SEQ ID NOS. 70 and 71.

These vectors were constructed in two steps. First, a fragment containing the CMV promoter and tri-partite leader sequence (TPL) of pAvCvIx (Fig. 26) was excised by digestion with MluI and BgIII, which cut at bp 473 and 1375 respectively. The restriction site overhangs were filled in with Klenow. Blunt ended DNA fragments containing the 6x2C7-SV40 or 6x2C7-TATA enhancer/promoter regions of the previously described reporter plasmids were ligated into this backbone to create pAV-6x2C7SV40 and pAV-6x2C7TATA shuttle plasmids. Next, DNA fragments containing the Luciferase or murine endostatin transgenes were ligated into the EcoRV site of the appropriate shuttle plasmids to create pAv6x2C7SV40-Luc (lox) or pAv6x2C7TATA-mEndo (lox).

30 Construction of a Right end vector plasmid

To complete the vector construction, a plasmid containing the remainder of the viral vector genome is required. This plasmid, referred to as pSQ3, which is shown in Fig. 27, contains a pBR322-derived backbone, ampicillin resistance gene and the adenovirus serotype 5 genome, beginning at Ad5 bp 3329, through the right ITR, with deletions in the E2a and E3 region as described previously (Gorziglia *et al.* (1996) *J. Virol.* 70:4173-4178). In addition, this plasmid has two important features, a loxP site inserted at the Bam HI site (bp 31569) just upstream of the Ad5 sequences, and a Cla I site at the end of the viral 5' ITR. This Cla I site is used to linearize the plasmid and expose the right ITR during vector construction.

Vector Assembly and propagation

Three adenoviral vectors encoding fusion protein regulators, Av3CV-C7LBDAS(G521R), Av3CV-C7LBDBS(G521R), and Av3CV-C7LBDBS(G400V) and two vectors containing regulatable transgenes, Av3SV-LUC and Av3TATA-Endo were constructed. Each vector was generated by a standard procedure. Briefly, for each vector construct, three plasmids, pSQ3 (pre digested with ClaI), the appropriate left end shuttle plasmid (e.g. pAvCv-C7LBD As(G521R), or pAv6X2C7SV40-Luc (lox), pre-digested with NotI and Afl II, and an expression plasmid for the Cre recombinase, pCMV-CRE, were cotransfected at a weight ratio of 3:1:1 into dexamethasone induced AE1-2a cells (Gorziglia *et al.*) using Promega's Profection Kit. About 1 week after transfection, cells were harvested and lysed by 4 cycles of freeze/thaw. The resulting cell lysate was passed onto fresh dexamethasone induced AE1-2a cells and the culture maintained about a week until cytopathic effect (CPE) was observed. This process was repeated several cycles until sufficient material was obtained to purify the vector by CsCl equilibrium density centrifugation. Once purified, vectors are quantitated by lysing in buffer containing 10mM Tris, 1mM EDTA, 0.1% SDS for 15 minutes at 56°C, cooling and reading the absorbance at 260 nm wavelength (OD260).

The OD260 reading is converted to a virus particle concentration using 1
OD260 unit = 1.1×10^{12} particles/ml.

Results

In Vitro Regulation with Adenovirus Vectors

5 The ability to regulate expression of a transgene delivered by an adenovirus vector was demonstrated by the following experiment. Hela cells were infected with a mixture of two adenovirus vectors, one containing a fusion protein regulator either (Av3-C7LBD-A(G521R) or Av3-C7LBD-B(G52R), the other containing the 6x2C7SV40-luc cassette.

10 To determine the optimal ratio of target vector to effector vector, two different doses of the transgene or target vector (50 or 250 viral particles per cell) at three different ratios of effector vector (50, 250, 750 particles per cell for each target dose) were tested. Twenty four hours after vector transduction, the cells were treated where appropriate with

15 100 nM 4-OH-tamoxifen. Following an additional 24 hrs incubation, the cells were lysed and assayed for luciferase activity. For the Av3CV-C7LBD A(G521R) vector, the data indicate relatively low levels of luc expression in the absence of 4-OHT, a strong 4-OHT dependent induction and a dose dependent increase in luc activity as more fusion protein

20 regulator vector is used. At the highest doses (750 particles per cell) of chimeric regulator vector tested, tamoxifen-specific induction of 460 to 560 fold over basal was achieved at target vector doses of 250 and 50 particles per cell, respectively.

25 The same experiment carried out using the LBD B version of the chimeric regulator; Av3CV-C7LBD B(G521R). For this vector, the fold induction and absolute luciferase activity were about two fold lower than obtained with the As-based regulator. These results are consistent with all the previous transient transfection experiments performed with plasmids. Notably, a first generation of Av3-chimeric regulator vectors constructed with the RSV promoter driving the expression of the C7LBD gene did not yield good transgene upregulation of the Av3SV40-Luc

vector. Apparently, the expression level from the weaker RSV promoter was not adequate to produce the necessary levels of fusion protein.

In Vivo Regulation with Adenovirus Vectors

To demonstrate the effectiveness of the C7LBD regulators to control transgene expression *in vivo*, a study was designed to evaluate three important variables: 1) the effectiveness of regulators containing either the G400V or G521R mutations, 2) the ratio of target and effector vector, and 3) the dose of 4-OHT. The importance of the G400V and G521R mutations are as follows. While the G521R mutation is selectively responsive to 4-OHT and is not affected by endogenous estrogen, it requires about a 10-fold higher drug concentration than the G400V mutation to achieve maximum activity. While the G400V is active at a lower dose of 4-OHT, it is also subject to induction by estrogen and could show higher basal activity *in vivo*.

Details of the animal study are as follow. On study day 1, C57Bl/6 male mice were given a total adenovirus vector dose of 2×10^{11} particles via tail vein injection. On day two blood samples were collected, then animals were injected i.p. with 200 ul of sunflower seed oil containing 5% DMSO and either no, 50 ug, or 500 ug of tamoxifen (Sigma # T56448). Blood samples were collected daily for three days following drug administration, and on study days 8 and 10. At the completion of the study, murine endostatin levels were determined by ELISA (Accucyte Kit, Cytimmune Sciences, Maryland).

The study groups included the following:

Negative Control – 2×10^{11} particles Av3Null, Ad vector with no transgene
Positive Control – 2×10^{11} particles Av3RSV-mEndo, constitutively expresses endostatin from the RSV promoter.
1:1 As521 – Received 1×10^{11} particles of Av3TATA-mEndo and 1×10^{11} particles of Av3Cv-C7LBDA(G521R); no treatment (basal) or + 50 μ g tamoxifen.

1:1 Bs400 – Received 1×10^{11} particles of Av3TATA-mEndo and 1×10^{11} particles of Av3Cv-C7LBDBs(G400V); no treatment (basal) or + 50 μ g tamoxifen.

In addition, groups 5 and 6 were similar to groups 3 and 4, but animals received 0.5×10^{11} of the Av3TATA-mEndo vector and 1.5×10^{11} of the C7LBD regulator vector, for a 1:3 ratio of target to effector. Groups 3 – 6 each contained no drug, 50 μ g, and 500 μ g tamoxifen treatment sub-groups.

The results showed a dramatic induction of murine endostatin following the day 2 administration of 50 μ g of tamoxifen. The highest level of induction was observed on day 3, the day immediately following drug administration. Compared to the basal level observed on day 3 in the no tamoxifen groups, the C7LBDA(G521R) and C7LBDBs(G400V) regulators gave comparable fold induction, approximately 17 fold, and comparable absolute levels of expression, around 1500 ng/ml. In this study, the endogenous murine endostatin levels in an untreated mouse cohort was 20 ± 7 ng/ml. The drug-induced endostatin expression rapidly declines by day 5, three days after drug administration, which is presumably due to the clearance of the tamoxifen and biological half life of the endostatin protein. In contrast, expression in the Av3RSV-mEndo treatment group persists at 200 ng/ml through day 15. In the 1:3 target to effector ratio groups, tamoxifen-induced expression reached 600 – 900 ng/ml, approximately 1/2 the level in the 1:1 ratio cohorts. This result indicates that *in vivo*, the transgene-containing vector, not the fusion protein-encoding vector, is limiting for absolute protein expression. Furthermore, endostatin expression in the animals treated with 500 μ g tamoxifen was comparable to the animals treated with only 50 μ g, indicating that the lower dose of tamoxifen is sufficient to fully activate the As(G521R) and Bs(G400V) regulators. Finally, the comparable low basal level of endostatin observed in the As(G521R) and Bs(G400V) groups suggests that the endogenous level of estrogen in the C57Bl/6

mice is not sufficient to induce the estrogen-responsive Bs(G400V) regulator. An elevation in basal endostatin levels observed at days 3 – 5 appeared to be a non-specific effect resulting from adenovirus vector administration, since the Av3Null vector has an effect similar to the 5 Av3TATA-mEndo containing groups.

Conclusions

The *in vitro* and *in vivo* results shown in this Example, demonstrate that the ZFP-LBD fusion proteins can be efficiently delivered via an adenovirus vector and can be expressed in sufficient amounts to provide 10 high levels of drug-dependent control of a transgene in animals.

Furthermore, the data show that the basal level of expression from the 6x2C7-minimal promoter constructs tested in an adenovirus vector give relatively low levels of expression, even when the fusion protein is expressed in the same cell. Thus, the system is highly drug dependent 15 and allows for substantial regulation of the vector-delivered transgene. Taken together, these data evidence the effectiveness of this system for gene therapy applications.

EXAMPLE 20

Construction and evaluation of the Cys₂-His₂ Zinc finger DBD-ERLBD regulators in Lentiviral Vectors

In order to demonstrate controlled gene expression in an integrated vector system, the the regulatory system described in Example 19 with the adenoviral vectors were used to develop a series of lentiviral vectors. These vectors contained either the ZFP-LBD fusion protein linked to the 20 immediate early CMV promoter or a regulatable transgene (either eGFP or luciferase) linked to the 6 X 2C7 array of C7 binding sites and either the minimal promoter from SV40 or C-fos TATA. The fusion protein-encoding vector and the regulatable transgene vector can then be used to generate lentiviral vector supernatant. The supernatant can be used to stably 25 transduced human cells either singly or in parallel. Stable cell lines containing the integrated vectors can then be induced with the 30

appropriate activating drug (e.g., 4-OH-tamoxifen) and gene expression is measured as fold induction in the presence and absence of drug.

Construction of Lentiviral Vectors encoding the ZFP-LBD fusion protein or the regulatable transgene.

5 The generation of lentiviral vectors and vector supernatant involves 3 main steps: first a gene or region of interest is inserted into shuttle vector backbone plasmid containing all of the viral cis-elements for transcription, packaging, reverse transcription, and integration. Second, the lentiviral vector shuttle plasmid is co-transfected into human 293 cells
10 along with plasmids providing the packaging functions (gag, pol, and env). Typically the transfections include 10 µg of vector plasmid, 10 µg of packaging plasmid and 1 µg envelope plasmid (Vesicular Stomatitis virus G envelope) using a Profection Calcium Phosphate transfection kit. Third, the culture supernatant containing the lentiviral vector is harvested
15 (between 24 and 48 hours post transfection) and used to transduce naïve human target cells.

Construction of HIV-1 based vectors

An HIV-1-based vector system containing an internal CMV promoter was constructed from an infectious HIV-1_{IIIB} provirus cDNA (pHIV-IIIB) The infectious proviral cDNA was generated by PCR from DNA isolated from H-9 cells chronically infected with HIV-1_{IIIB}. The gag/pol and env sequences of pHIVIIIB were removed by digestion and excision of a PstI-KpnI fragment. Replacing the gag/pol and env sequences was a PstI/Kpn polylinker containing unique multiple cloning sites to form the intermediate vector p2XLTR. The Rev response element (RRE) fragment from HIVIIIB, required for proper vector RNA processing, was inserted downstream of the truncated gag sequences of p2XTR to form the construct pHIVec. An Asel-Xba CMV-eGFP reporter fragment derived from pEGFP-N1 (Clontech, Palo Alto, CA) was cloned into the NdeI-Xba site of pHIVec to generate pHIVCMVGFP. pHIVCMV-X was

generated by removal of the eGFP fragment by KpnI digestion and religation.

Construction of pHIVCMV-C7LBD/A(G521R)

The AS521R (C7LBD/A(G521R) coding fragment derived from C7LBDAS by digestion with NotI, T4 DNA polymerase fill-in, and EcoRI site was cloned into pHIVCMV-X cloned downstream of the CMV promoter into a EcoRI/SmaI restriction site. As a control for induction, an HIV vector containing a constitutive transactivator and DBD chimera was generated, pHIVCMV-C7VP16. A HindIII-NotI restriction fragment from pCDNA3-C7VP16 containing the C7VP16 coding fragment was inserted downstream of the CMV promoter at the Sma site of pHIVecCMV-X.

Construction of pHIV6X2C7Sv and pHIV6X2C7TATA luciferase vectors

A BamHI-XbaI restriction fragment containing the 6X2C7TATA luciferase fragment was isolated from pTATA6X2C7Luc and cloned downstream of the RRE at the SpeI-XbaI restriction sites. A MluI-BstBI restriction fragment containing the 6X2C7Sv luciferase fragment was isolated from pGL3-6X2C7SvLuc and cloned downstream of the RRE at the SpeI-XbaI restriction sites.

Evaluation of the ZFP-LBD fusion proteins and regulatable lentiviral vectors

Transduction of HeLa cells by inducible lentiviral vectors

Subconfluent HeLa cells were transduced with either HIV6X2C7SvLuc or HIV6X2C7TATALuc vector supernatant for 24 hours followed by transduction with HIVAS521R lentiviral vector supernatant. Cells were allowed to recover from infection for 24 hours in fresh culture medium after which 4-OH-tamoxifen (100 or 1000 nM) was added to the culture for an additional 24 hours. Cells were lysed in a standard luciferase lysis buffer, subjected to freeze thaw and analyzed for luciferase activity using a luciferase assay kit (Promega). The results showed that cells infected with either HIV6X2C7SvLuc or HIV6X2C7TATALuc followed by transduction with HIVCMVAS521R

resulted in a 13.1 and 11.7 fold stimulation in luciferase activity respectively, when given 4-OH-tamoxifen.

Lentiviral Transduction of lentiviral integrated target vector populations

5 HeLa cells that had been previously transduced with either HIV6X2C7SvLuc or HIV6X2C7TATA⁺Luc were carried in culture for 9 passages without exposure to any ZFP-LBD fusion protein. On passage 10, cells were transduced with HIVCMVAS521R for 24 hours followed by the addition of 100 nm tamoxifen for an additional 24 hours. The
10 results show that HeLa cell lines containing an integrated HIV6X2C7SvLuc or HIV6X2C7TATA⁺Luc vector can be induced for luciferase expression by transduction of a LV containing AS521R + tamoxifen 31.4- and 22.5-fold, respectively.

15 These data demonstrate the effectiveness of the C2H2-LBD regulator for controlling expression of a transgene that is stably integrated into the host cell chromosome.

20 Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

WHAT IS CLAIMED IS:

1. A fusion protein, comprising a nucleotide binding domain operatively linked to a ligand binding domain derived from an intracellular receptor, wherein:

5 the nucleotide binding domain is a polydactyl zinc-finger peptide or modular portion thereof that specifically interacts with a contiguous nucleotide sequence of at least about 3 nucleotides; and

the fusion protein is a ligand activated transcriptional regulator.

10 2. The fusion protein of claim 1, further comprising an operatively linked transcription regulating domain.

3. The fusion protein of claim 1, wherein the intracellular receptor is a nuclear hormone receptor.

15 4. The fusion protein of claim 3, wherein the ligand binding domain derived from a nuclear hormone receptor has been modified to change its ligand selectivity compared to the native hormone receptor.

5. The fusion protein of claim 4, wherein the modified ligand-binding domain is not substantially activated by endogenous ligands.

20 6. The fusion protein of claim 1, wherein zinc-finger peptide binds to a sequence of nucleotides of the formula $(GNN)_n$, where G is guanidine, N is any nucleotide and n is an integer from 1 to 6.

7. The fusion protein of claim 6, wherein n is 3 to 6.

25 8. The fusion protein of claim 1, wherein the zinc-finger peptide is comprised of modular units from a C2H2 zinc-finger peptide or a variant thereof that specifically interacts with a sequence of nucleotides and targets the fusion protein to a exogenous or endogenous gene that comprises the sequence of nucleotides.

9. The fusion protein of claim 1, wherein the zinc finger peptide is comprised of at least one zinc finger or a variant thereof that specifically binds to a targeted nucleic acid molecule.

30 10. The fusion protein of claim 9, that comprises at least three zinc fingers or variants thereof.

11. The fusion protein of claim 1, wherein the intracellular receptor is a nuclear hormone receptor selected from the group consisting of estrogen receptors, progesterone receptors, glucocorticoid- α receptors, glucocorticoid- β receptors, mineralocorticoid receptors, androgen receptors, thyroid hormone receptors, retinoic acid receptors, retinoid X receptors, Vitamin D receptors, COUP-TF receptors, ecdysone receptors, Nurr-1 receptors and orphan receptors.

5 12. The fusion protein of claim 1, wherein the intracellular receptor is a steroid receptor.

10 13. The fusion protein of claim 4, wherein the hormone receptor is a progesterone receptor variant or an estrogen receptor variant, wherein a receptor variant comprises a ligand binding domain that has selectivity and sensitivity for endogenous and exogenous ligands that differ from its native ligands.

15 14. The fusion protein of claim 2, wherein the transcription regulating domain comprises a transcription activation domain.

20 15. The fusion protein of claim 2, wherein the transcription regulating domain comprises a transcription activation domain selected from the group consisting of VP16, VP64, TA2, STAT-6, p65 and derivatives, multimers and combinations thereof that have transcription activation activity.

25 16. The fusion protein of claim 14, wherein the transcription regulating domain comprises a nuclear hormone receptor transcription activation domain or variant thereof that has transcription activation activity.

17. The fusion protein of claim 14, wherein the transcription regulating domain comprises a steroid hormone receptor transcription activation domain or variant thereof.

30 18. The fusion protein of claim 14, wherein the transcription regulating domain comprises a viral transcription activation domain or variant thereof that has transcription activation activity

19. The fusion protein of claim 18, wherein the transcription regulating domain comprises a VP16 transcription activation domain or variant thereof.

5 20. The fusion protein of claim 2, wherein the transcription regulating domain comprises a transcription repression domain.

10 21. The fusion protein of claim 20, wherein the transcription repression domain is selected from the group consisting of ERD, KRAB, SID, Deacetylase, and derivatives, multimers and combinations thereof such as KRAB-ERD, SID-ERD, (KRAB)₂, (KRAB)₃, KRAB-A, (KRAB-A)₂, (SID)₂, (KRAB-A)-SID and SID-(KRAB-A).

22. The fusion protein of claim 2 encoded by the sequence of nucleotides set forth in any of SEQ ID Nos. 1-18.

15 23. A nucleic acid molecule, comprising a sequence of nucleotides encoding the fusion protein of claim 1.

24. A nucleic acid molecule, comprising a sequence of nucleotides encoding the fusion protein of claim 2.

25. The nucleic acid molecule of claim 23, wherein the the fusion protein is encoded by a sequence of nucleotides set forth in any of SEQ ID Nos. 1-18.

20 26. A vector, comprising a sequence of nucleotides encoding the fusion protein of claim 1.

27. A vector, comprising a sequence of nucleotides encoding the fusion protein of claim 2.

28. A cell, comprising the expression vector of claim 26.

25 29. A cell, comprising the expression vector of claim 27.

30. The cell of claim 28 that is a eukaryotic cell.

31. The cell of claim 29 that is a eukaryotic cell.

32. The vector of claim 26 that is a viral vector.

33. The vector of claim 27 that is a viral vector.

30 34. The vector of claim 32, wherein the viral vector derived from a DNA virus or a retrovirus.

35. The vector of claim 34 that is selected from the group consisting of an adenoviral vector, and adeno-associated viral vector, a herpes virus vector, a vaccinia virus vector and a lentiviral vector.

36. The vector of claim 33 that is a viral vector.

5 37. The vector of claim 36, wherein the viral vector derived from a DNA virus or a retrovirus.

38. The vector of claim 37 that is selected from the group consisting of an adenoviral vector, and adeno-associated viral vector, a herpes virus vector, a vaccinia virus vector and a lentiviral vector.

10 39. A combination, comprising:

a fusion protein of claim 1 or a nucleic acid molecule comprising a sequence of nucleotides that encodes the fusion protein; and

15 a regulatable expression cassette that comprises at least one response element recognized by the nucleic acid binding domain of the fusion protein.

40. The combination of claim 39, wherein the cassette comprises a gene that encodes a therapeutic product.

20 41. The combination of claim 39 that comprises a single composition that contains the fusion protein or nucleic acid molecule that encodes the fusion protein, and the regulatable expression cassette in a pharmaceutically acceptable excipient.

25 42. The combination of claim 39, wherein the fusion protein or nucleic acid molecule comprising a sequence of nucleotides that encodes the fusion protein, and the regulatable expression cassette are in separate compositions.

30 43. A composition for regulating gene expression comprising:
an effective amount of the fusion protein of claim 1 or a nucleic acid molecule comprising a sequence of nucleotides that encodes the fusion protein; and
a pharmaceutically acceptable excipient.

44. The composition of claim of claim 43 that is formulated for single dosage administration.

45. A composition for regulating gene expression comprising: an effective amount of the fusion protein of claim 2; and
5 a pharmaceutically acceptable excipient.

46. The combination of claim 39, wherein the regulatable expression cassette comprises 3 to 6 response elements.

47. A method for regulating gene expression in a cell, comprising:

10 introducing into a cell a fusion protein of claim 1 or a nucleic acid molecule that comprises a sequence of nucleotides that encodes the fusion protein; and

15 contacting the cell with a ligand that interacts with the binding domain in the fusion protein, whereby the fusion protein interacts with a target nucleic acid molecule to activate or repress transcription of a gene encoded by the fusion protein.

48. The method of claim 47, wherein the ligand binding domain is modified whereby it interacts with a non-natural ligand.

49. The method of claim 47, wherein the target nucleic acid
20 molecule is endogenous to the cell.

50. The method of claim 47, wherein the target nucleic acid molecule is introduced to the cell as part of an expression cassette.

51. The method of claim 47, wherein the expression cassette and fusion protein or nucleic acid encoding the fusion protein are
25 introduced at the same time.

52. The method of claim 47, wherein the expression cassette and fusion protein or nucleic acid encoding the fusion protein are introduced sequentially.

53. The method of claim 47, wherein ligand is delivered to the
30 cell after the fusion protein or nucleic acid molecule encoding the fusion protein is introduced into the cell.

54. The method of claim 47, wherein the nucleic acid molecule encoding the fusion protein comprises a vector.
55. The method of claim 54, wherein the vector is a viral vector.
56. The method of claim 47, wherein the cell is in a mammal.
57. The method of claim 50, wherein the expression cassette is contained in a vector.
58. The method of claim 57, wherein the vector is a viral vector.
59. The method of claim 50, wherein the cell is in a mammal.
60. The method of claim 47, wherein the ligand binding domain derived from a nuclear hormone receptor has been modified to change its ligand selectivity compared to the native hormone receptor.
- 10 61. The method of claim 60, wherein the modified ligand-binding domain is not substantially activated by endogenous ligands.
- 15 62. The method of claim 47, wherein zinc-finger peptide binds to a sequence of nucleotides of the formula (GNN)_n, where G is guanidine, N is any nucleotide and n is an integer from 1 to 6.
- 16 63. The method of claim 62, wherein n is 3 to 6.
- 17 64. The method of claim 47, wherein the zinc-finger peptide is comprised of modular units from a C2H2 zinc-finger peptide or a variant thereof that specifically interacts with a sequence of nucleotides and targets the fusion protein to a exogenous or endogenous gene that comprises the sequence of nucleotides.
- 20 65. The method of claim 47, wherein the zinc finger peptide is comprised of at least one zinc finger or a variant thereof that specifically binds to a targeted nucleic acid molecule.
- 25 66. The method of claim 65, that comprises at least three zinc fingers or variants thereof.
67. The method of claim 47, wherein the intracellular receptor is a nuclear hormone receptor selected from the group consisting of estrogen receptors, progesterone receptors, glucocorticoid- α receptors, glucocorticoid- β receptors, mineralocorticoid receptors, androgen

receptors, thyroid hormone receptors, retinoic acid receptors, retinoid X receptors, Vitamin D receptors, COUP-TF receptors, ecdysone receptors, Nurr-1 receptors and orphan receptors.

5 68. The method of claim 47, wherein the intracellular receptor is a steroid receptor.

69. The fusion protein of claim 1, wherein the polydactyl zinc-finger peptide or modular portion thereof specifically interacts with a contiguous nucleotide sequence of at least about 3 nucleotides to about 18 nucleotides.

10 70. A non-viral delivery system, comprising the fusion protein of claim 1 or a nucleic acid molecule encoding the fusion protein.

15 71. The non-viral delivery system of claim 70, further comprising a nucleic acid molecule that comprises an expression cassette containing a sequence of nucleotides with which the nucleic acid binding domain of the fusion protein interacts.

20 72. The non-viral delivery system of claim 70, wherein the non-viral delivery system is selected from the group consisting of DNA-ligand complexes, adenovirus-ligand-DNA complexes, direct injection of DNA, CaPO₄ precipitation, gene gun techniques, electroporation, liposomes and lipofection.

73. The fusion protein of claim 9, wherein the zinc finger peptide comprised of at least one zinc finger or a variant thereof specifically binds to a targeted nucleic acid molecule with a dissociation constant of less than about 1.0 nanomolar.

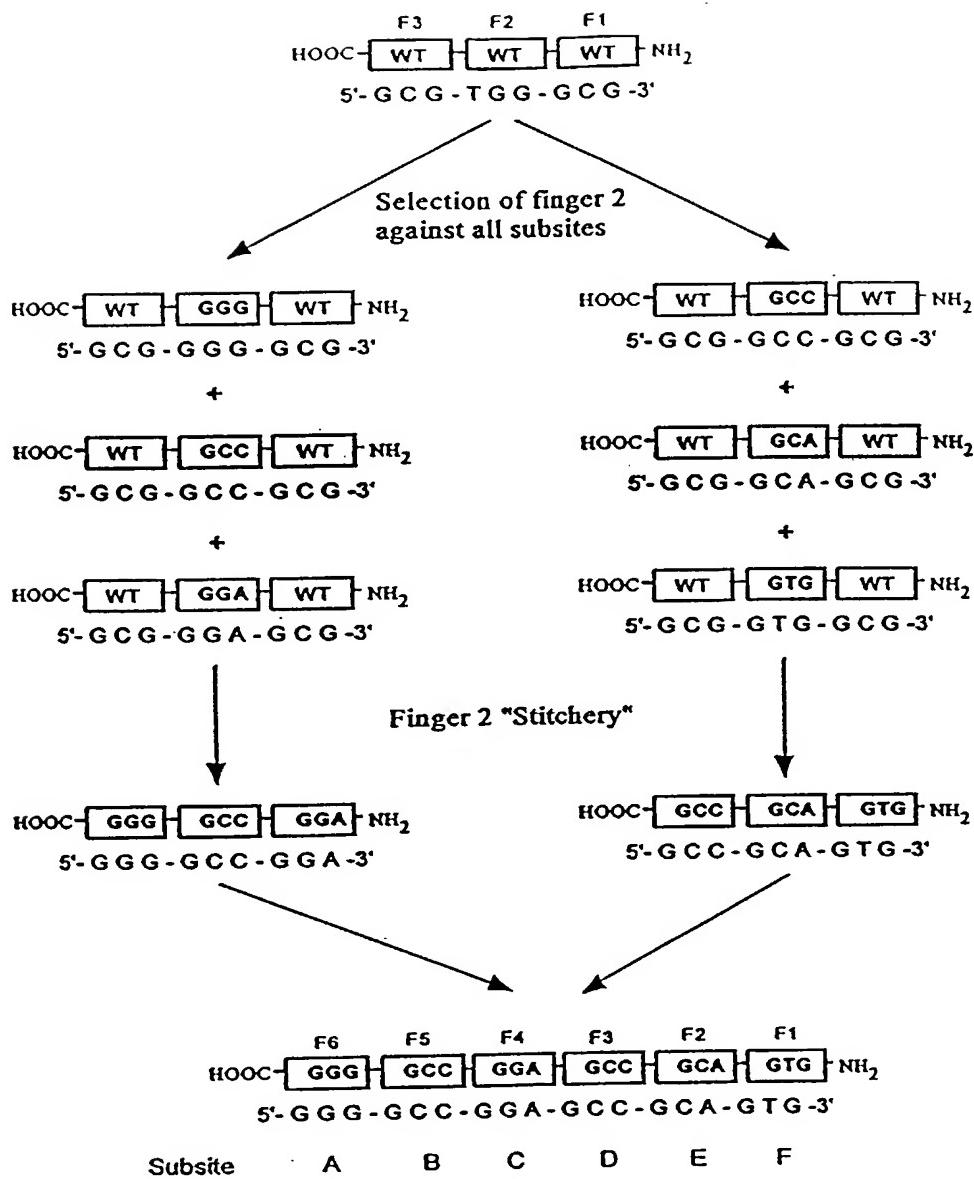


FIGURE 1

**Domains of a Steroid Nuclear Receptor
ex. Human Estrogen Receptor (ER)**

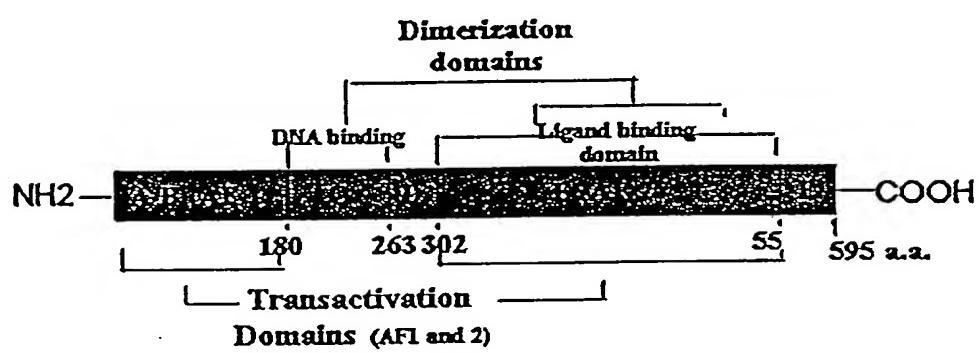
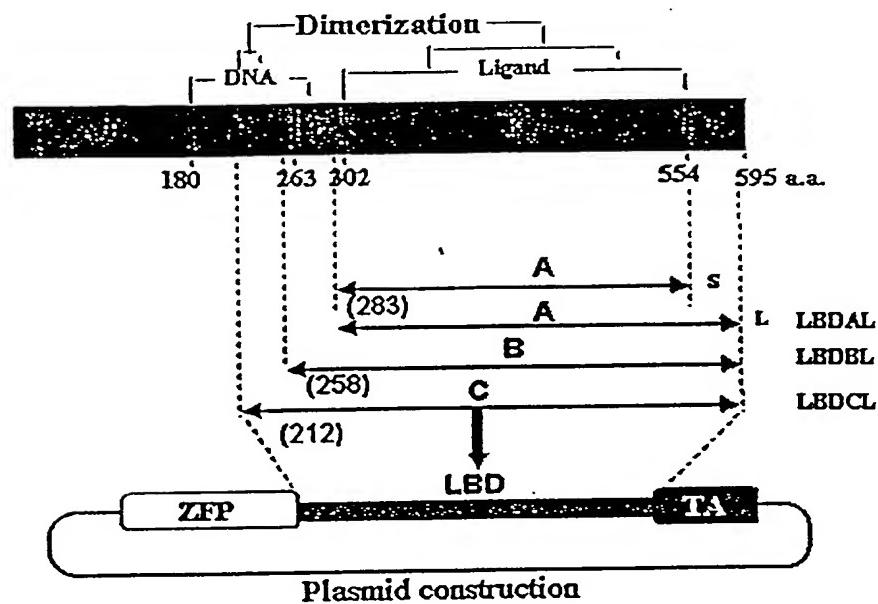


FIGURE 2

Cloning Strategy for DNA binding domain Replacement**FIGURE 3**

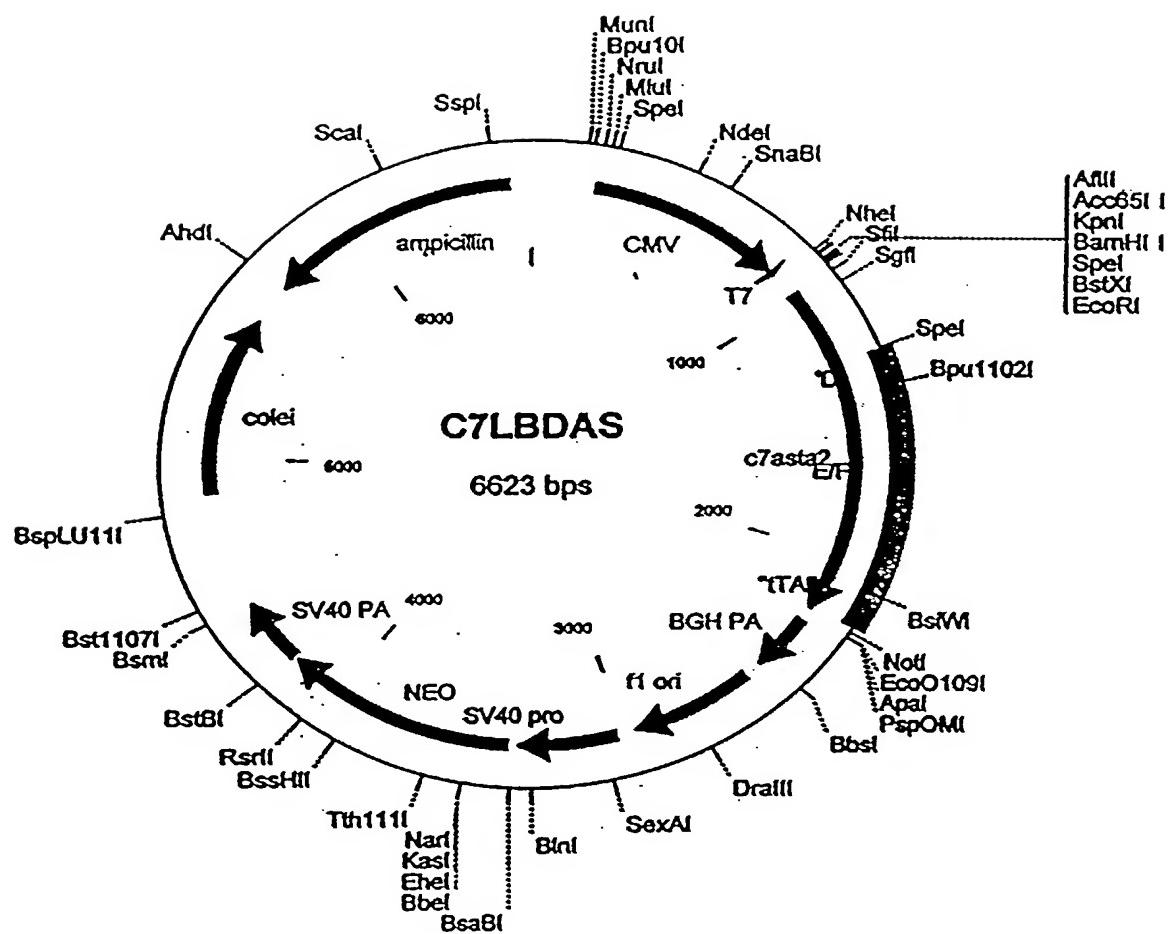
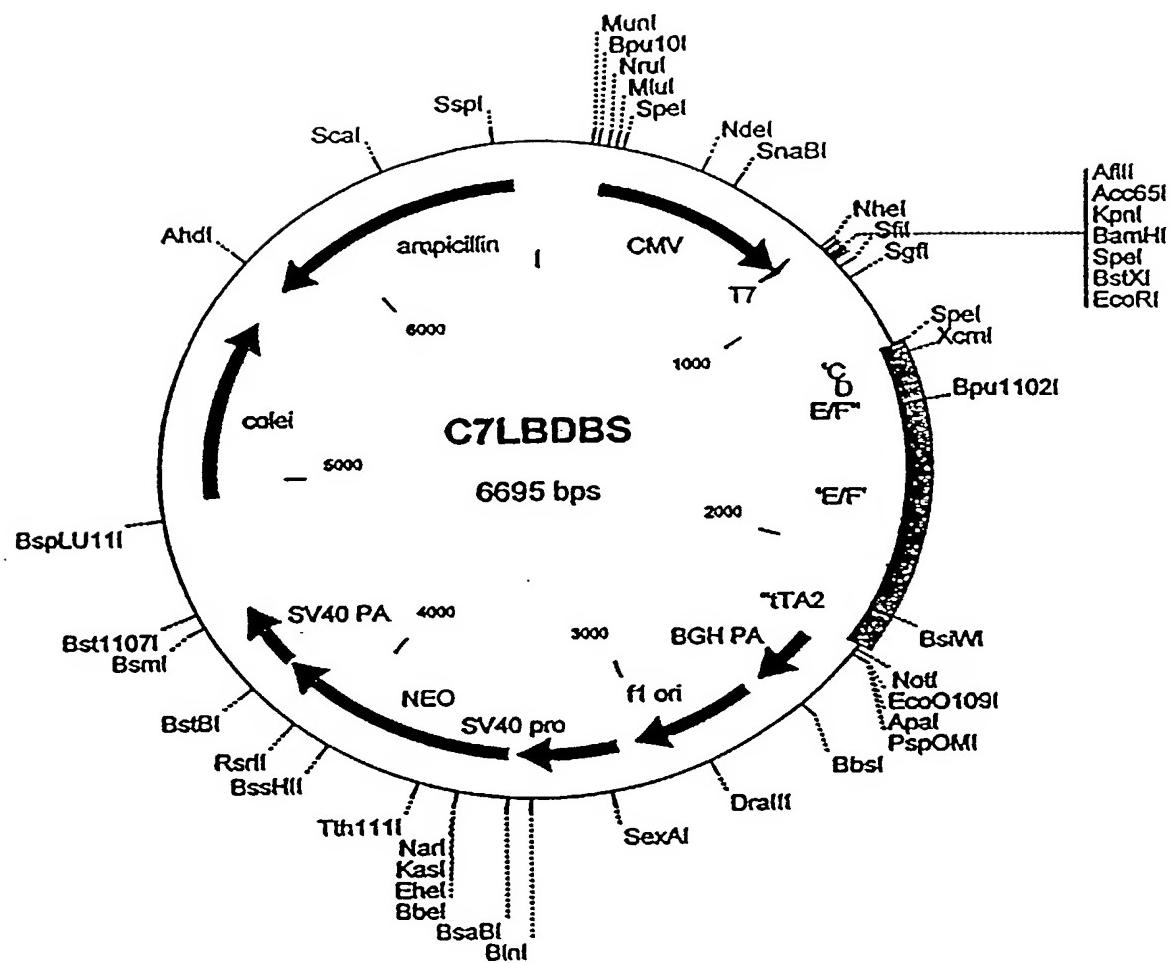
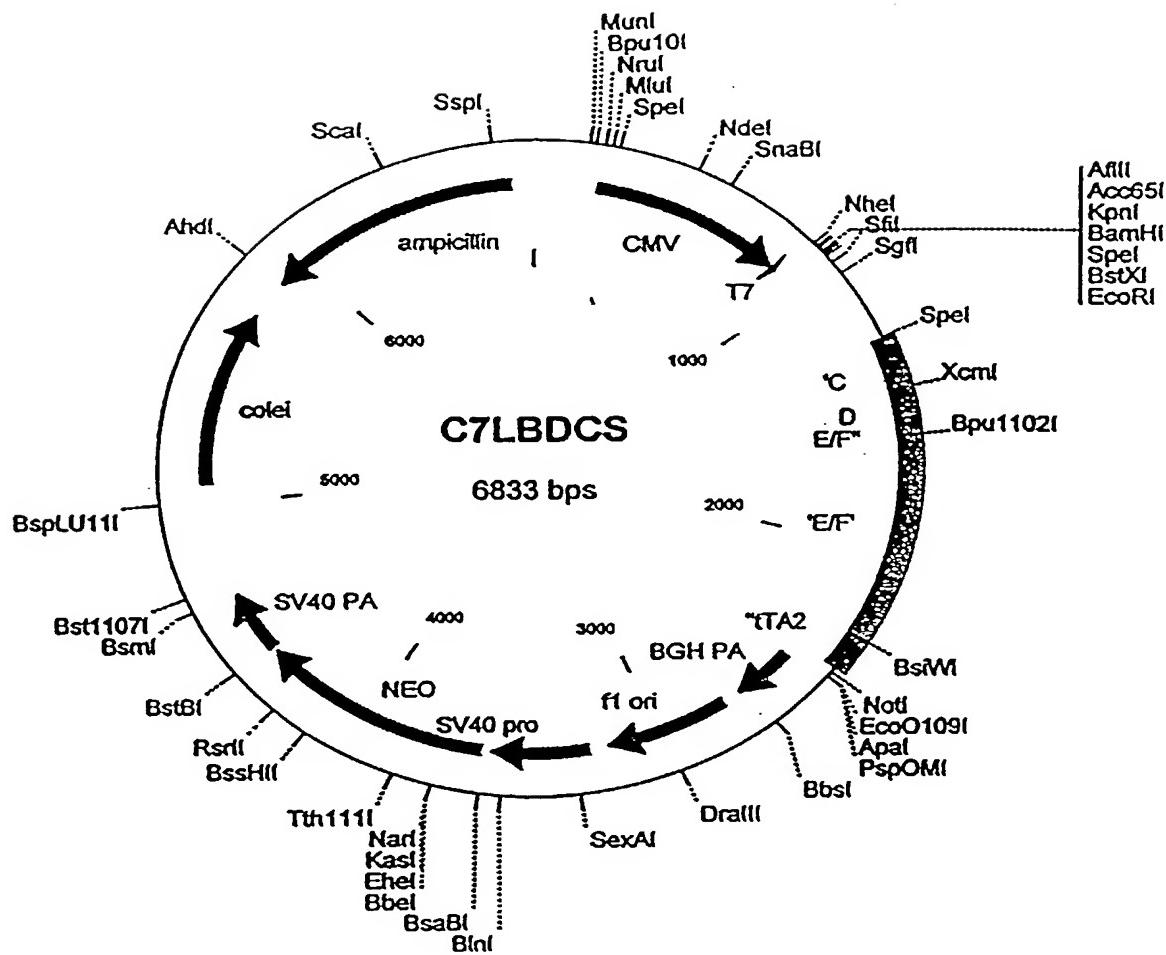


FIGURE 4

**FIGURE 5**

**FIGURE 6**

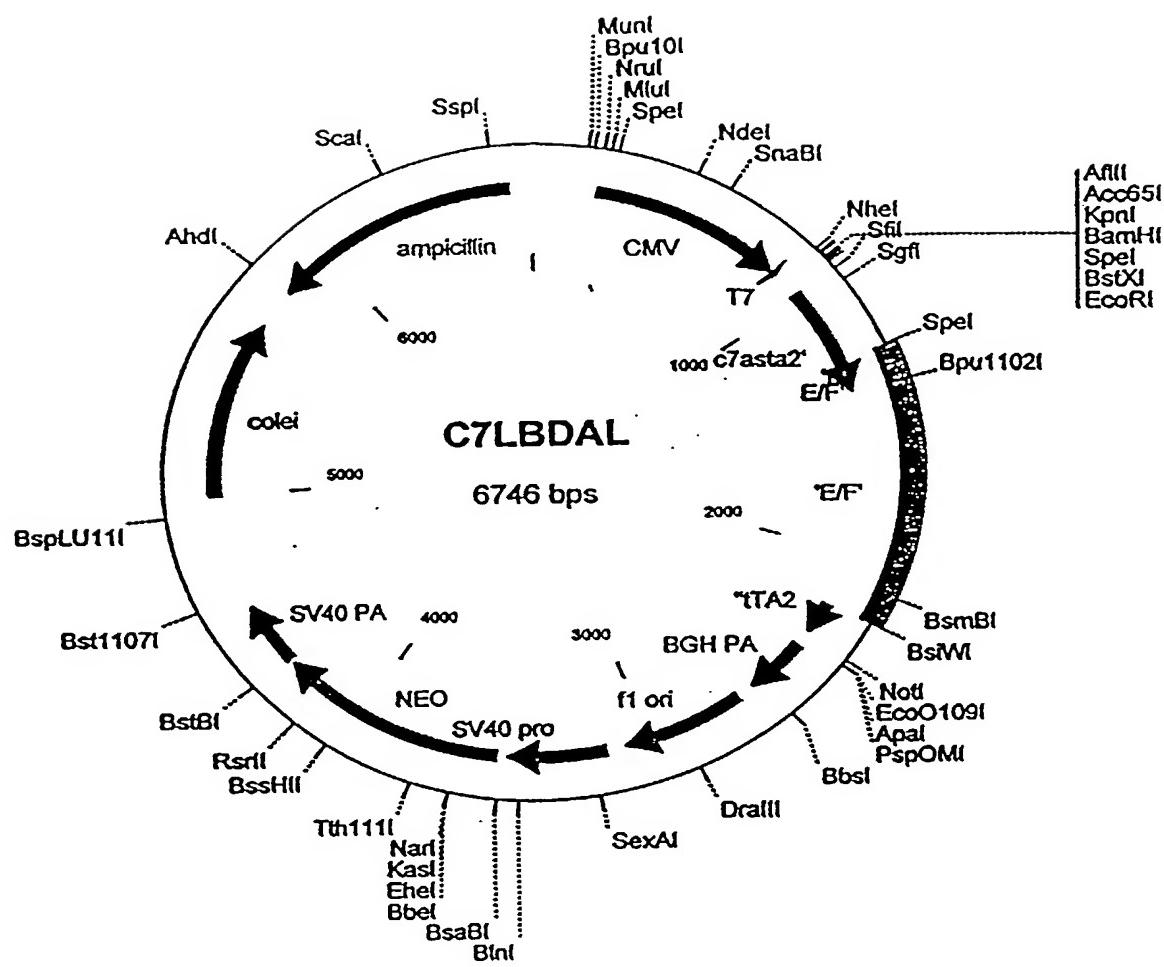
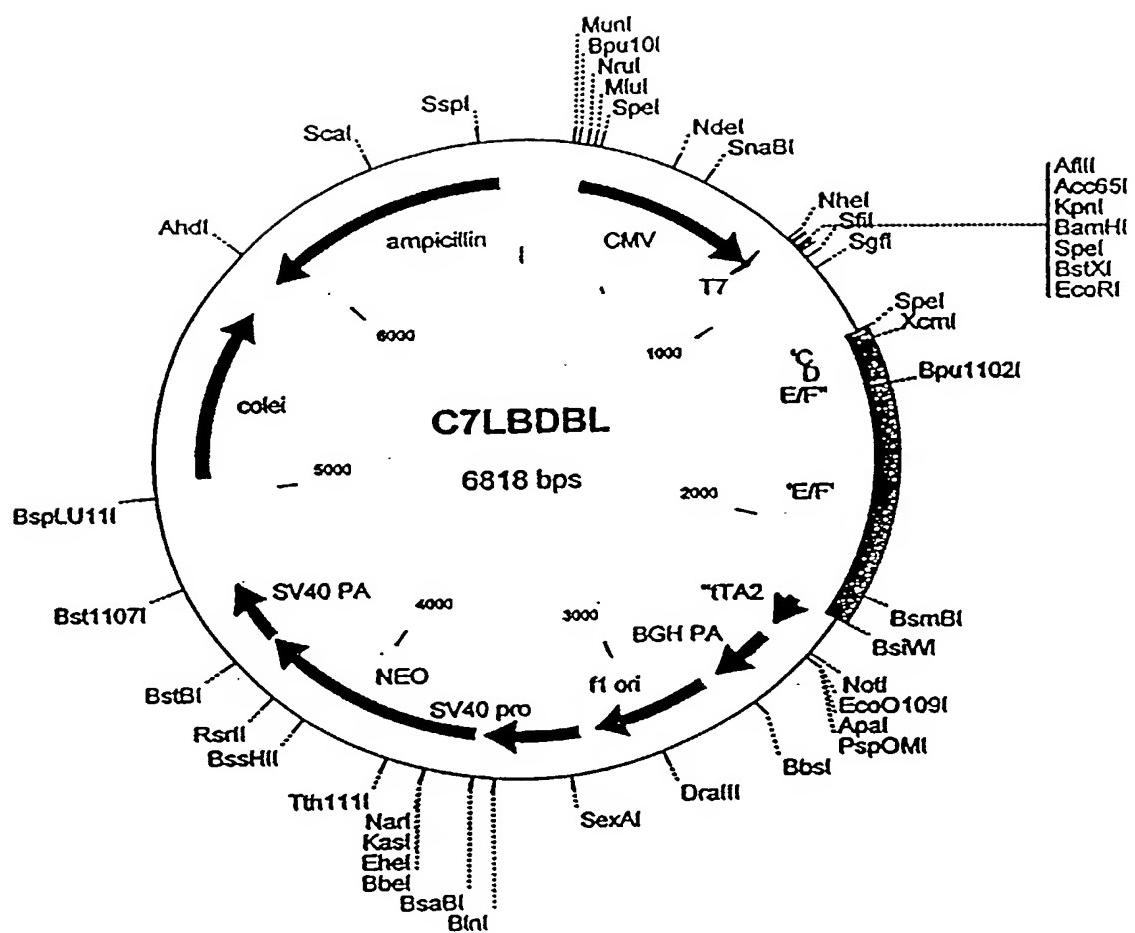
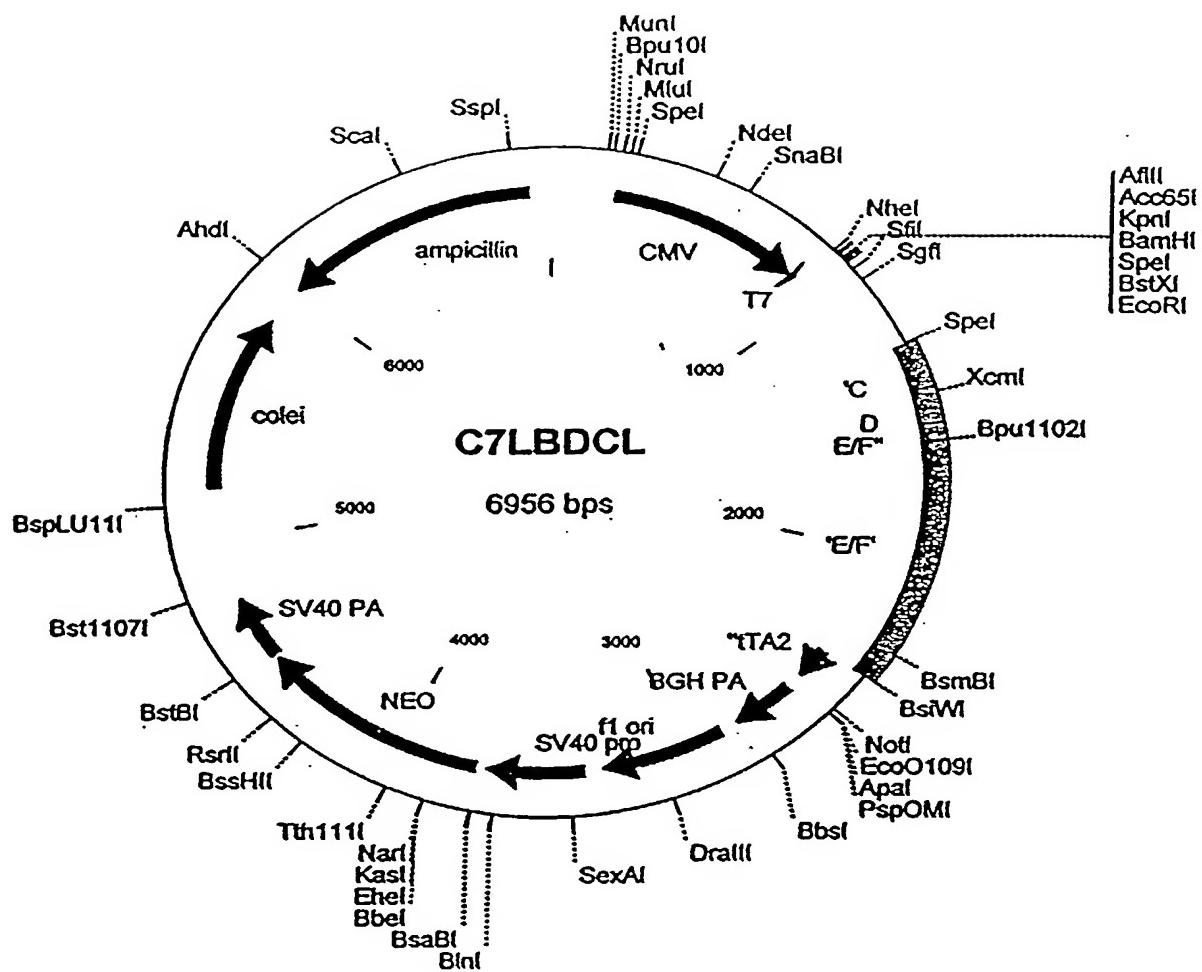


FIGURE 7

**FIGURE 8**

**FIGURE 9**

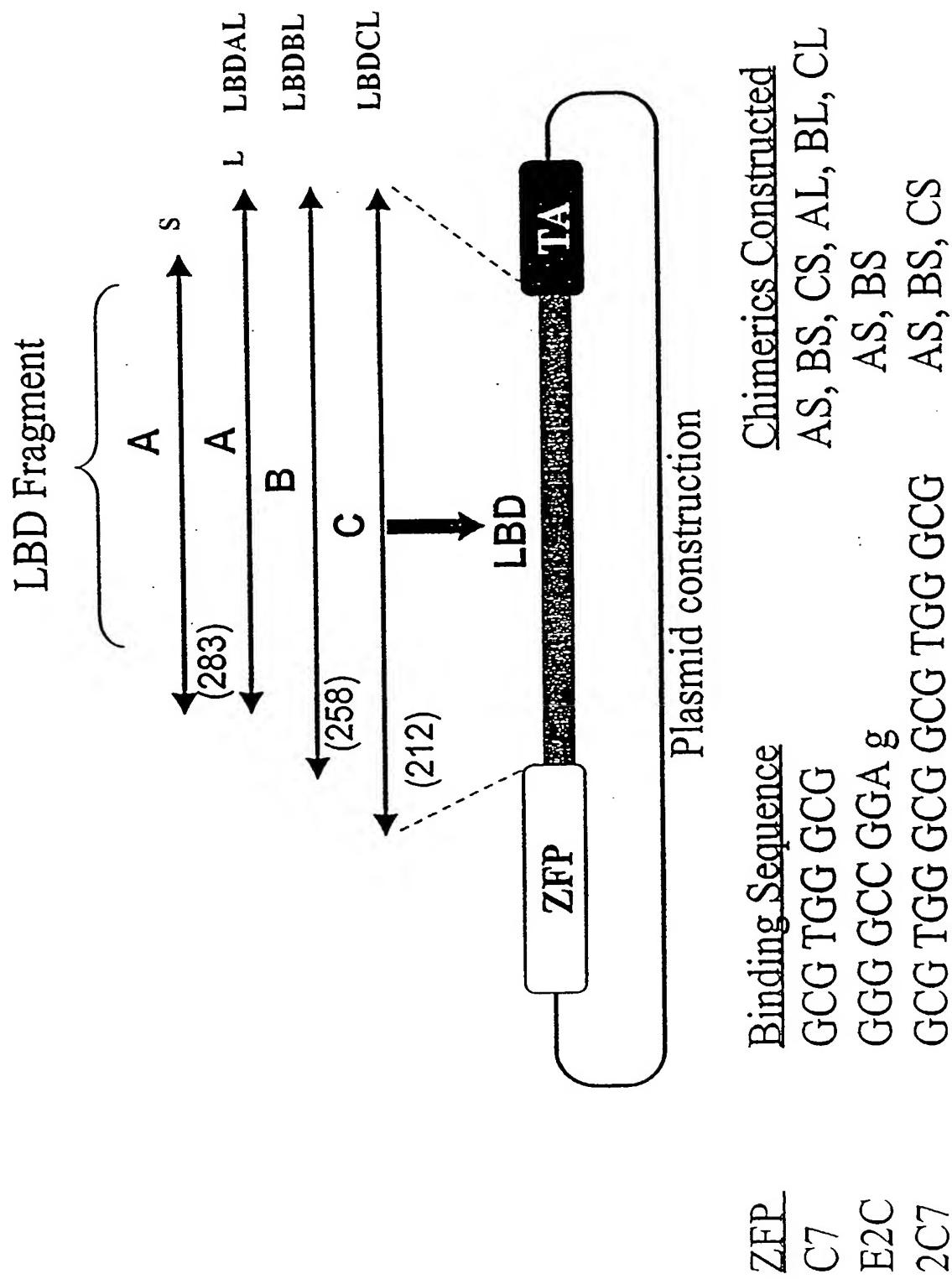


FIGURE 10

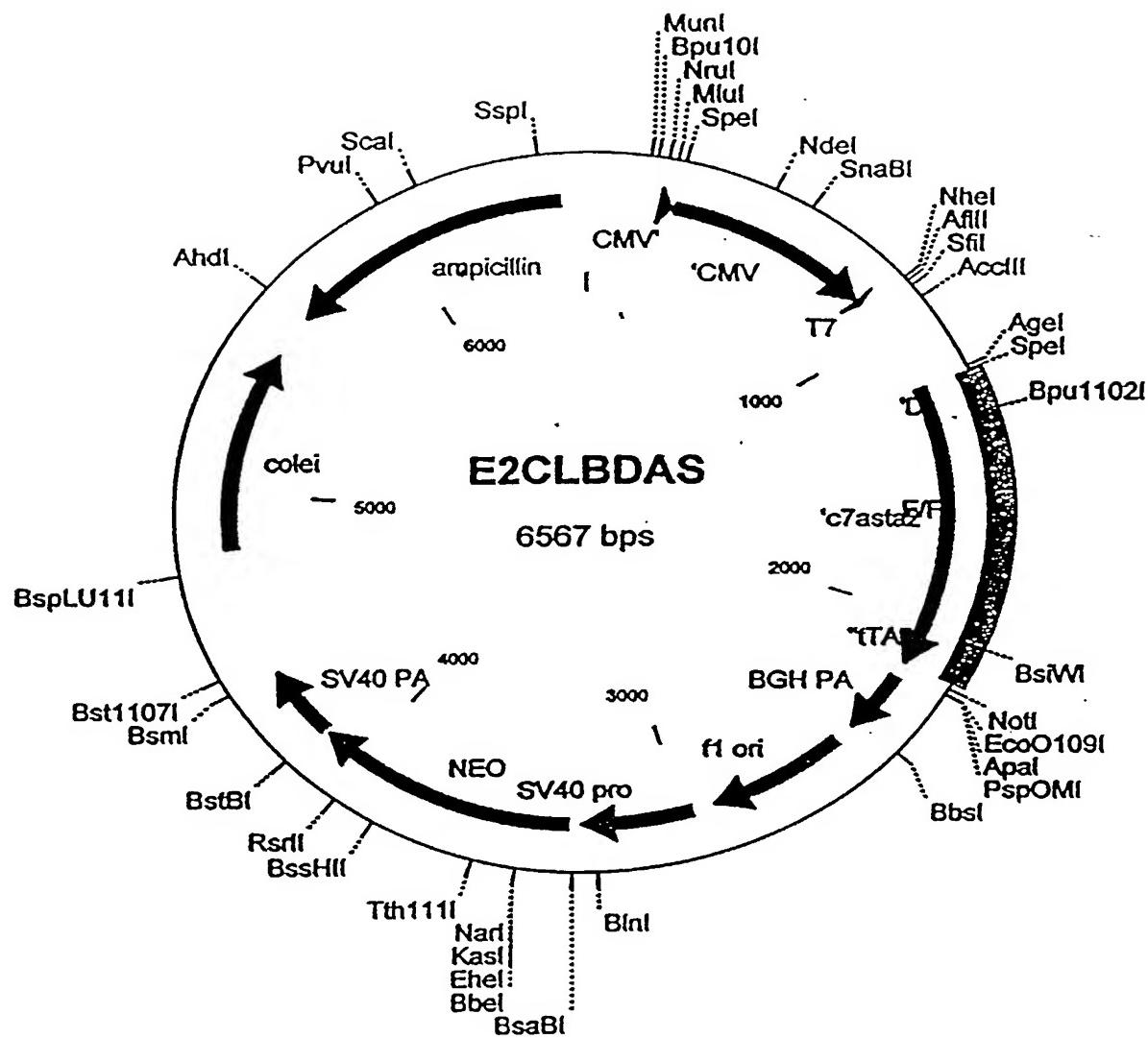


FIGURE 11

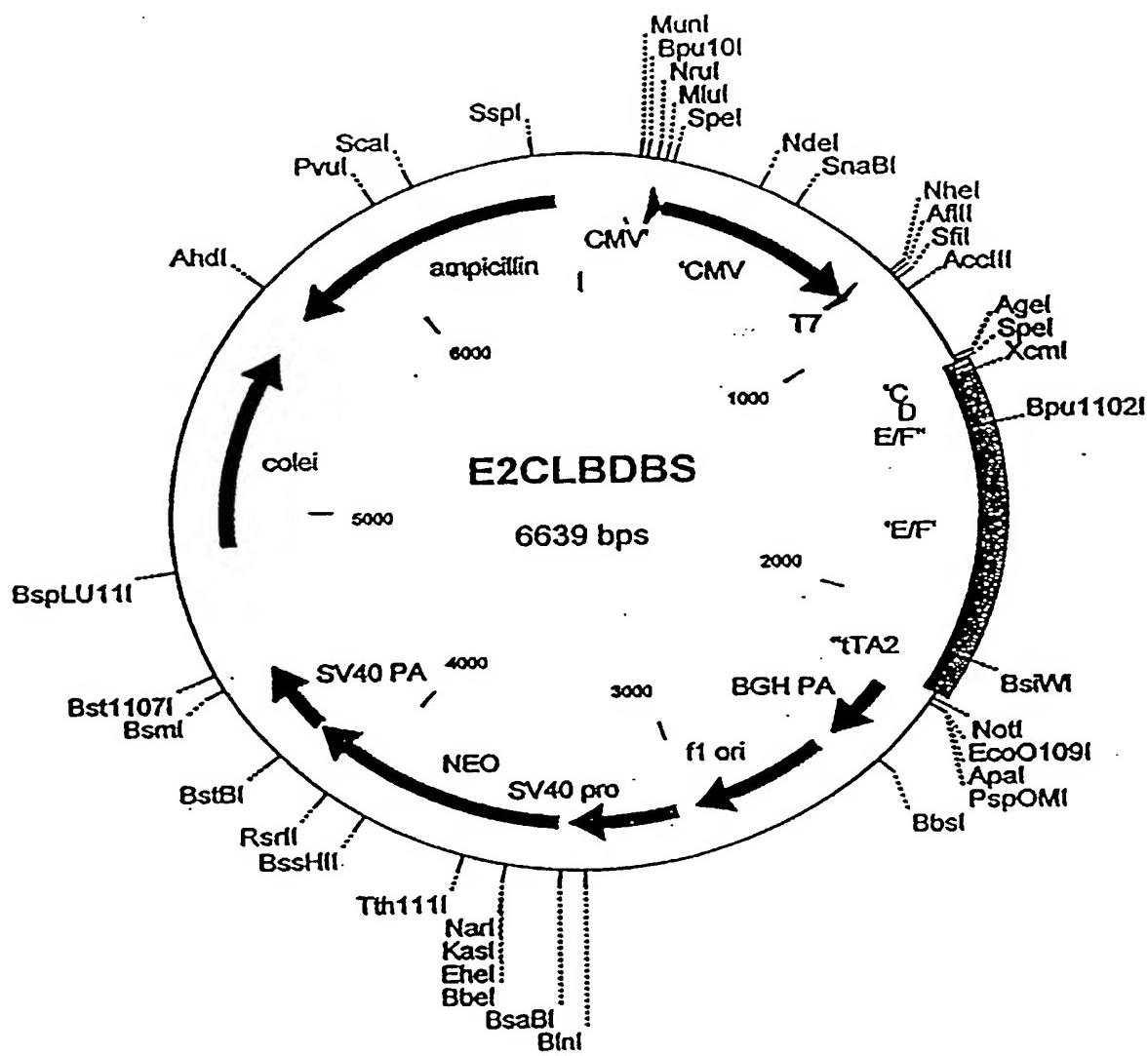
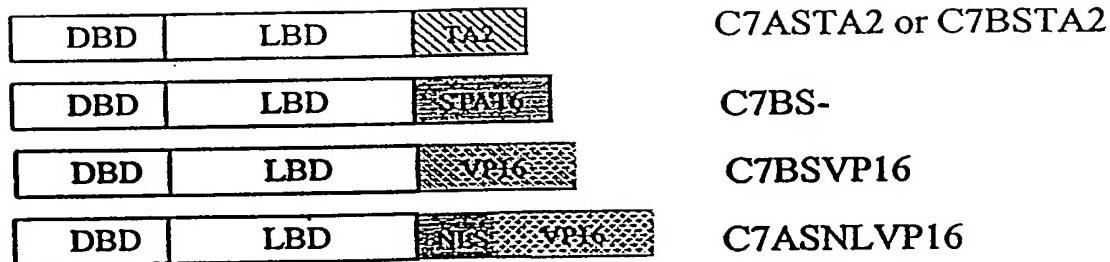
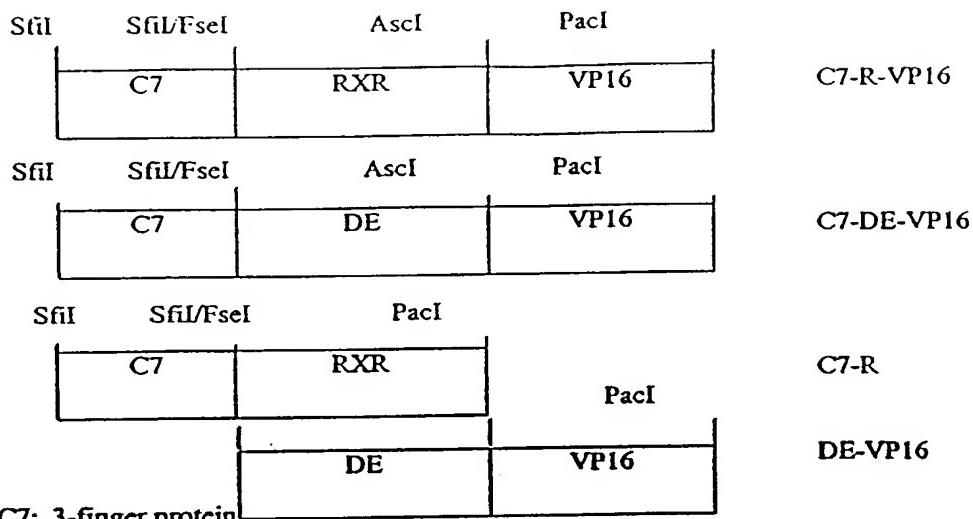


FIGURE 12

**FIGURE 13**



C7: 3-finger protein

RXR: mammalian RXR ligand binding domain

DE: Ecr ligand binding domain from *Drosophila*

BE: Ecr ligand binding domain from *Bombyx mori*

FIGURE 14

Subcloning for 2C7-LBD expression constructs

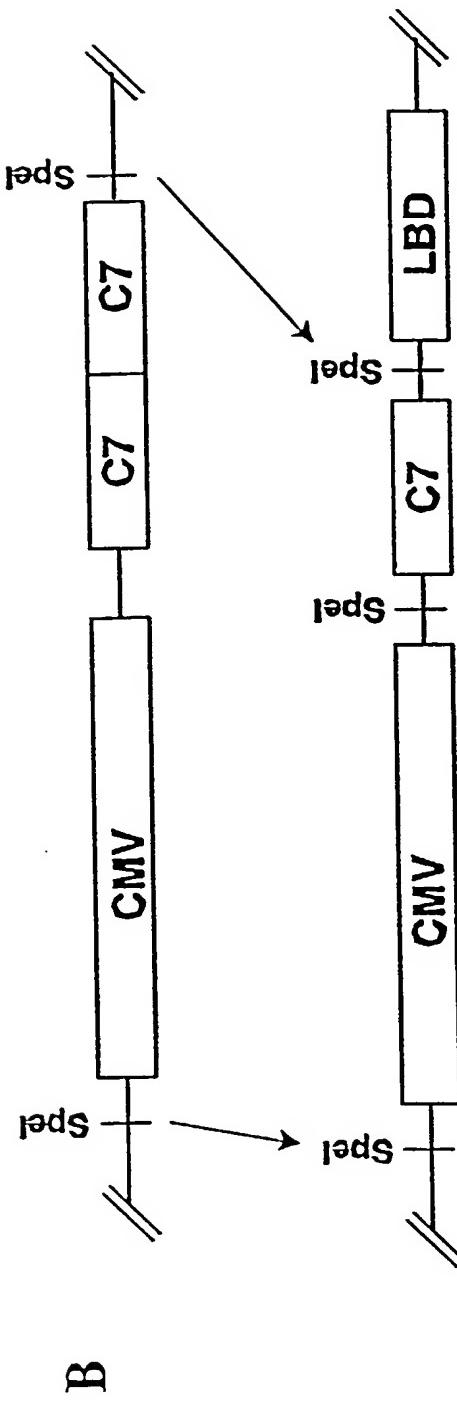
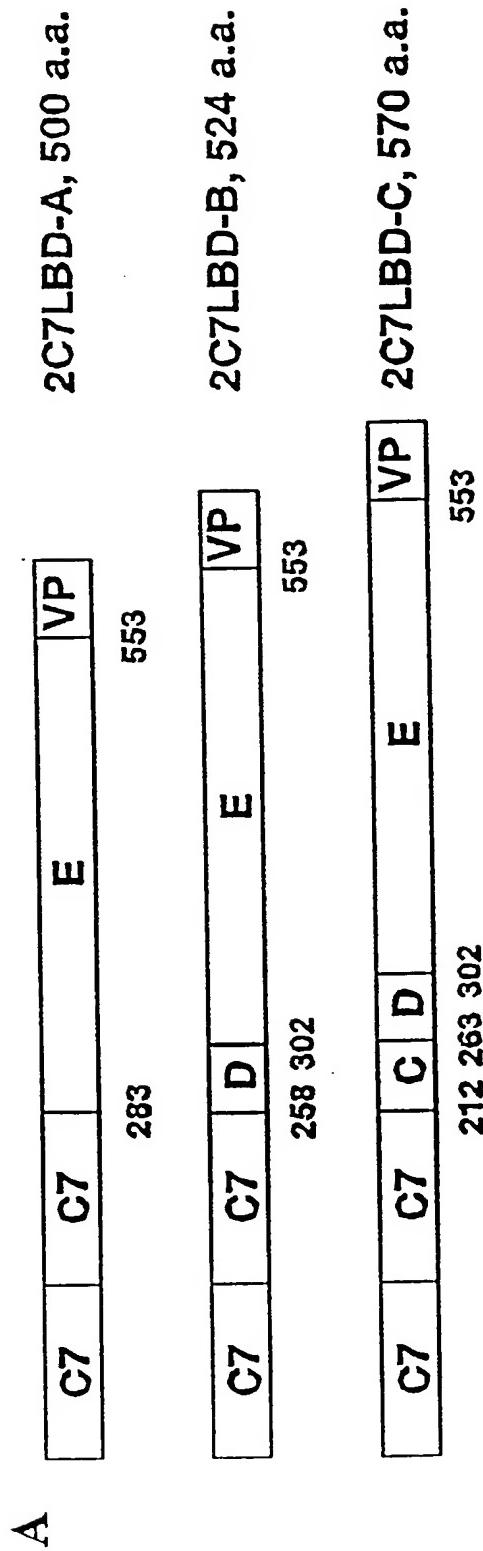


FIGURE 15

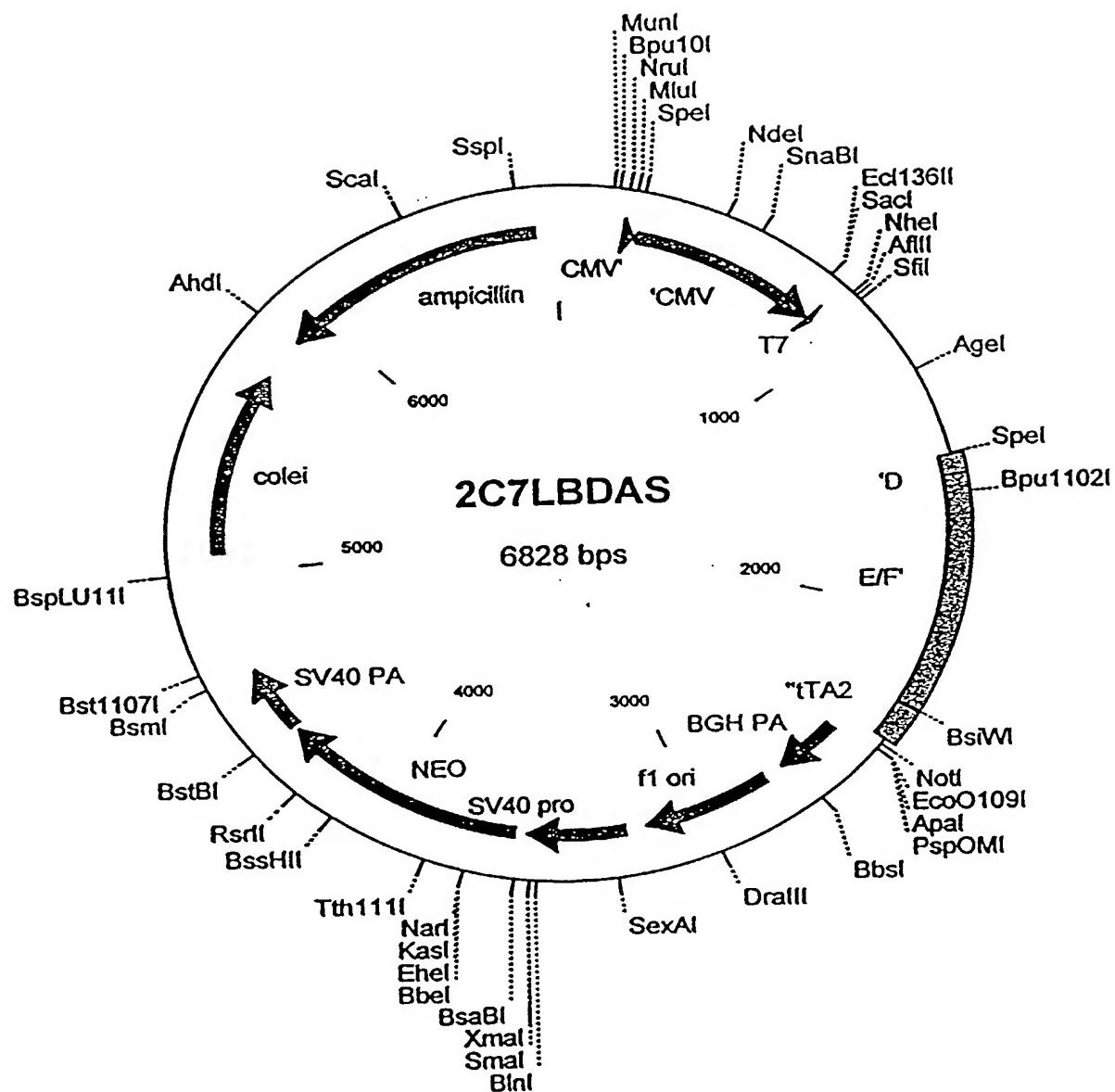


FIGURE 16

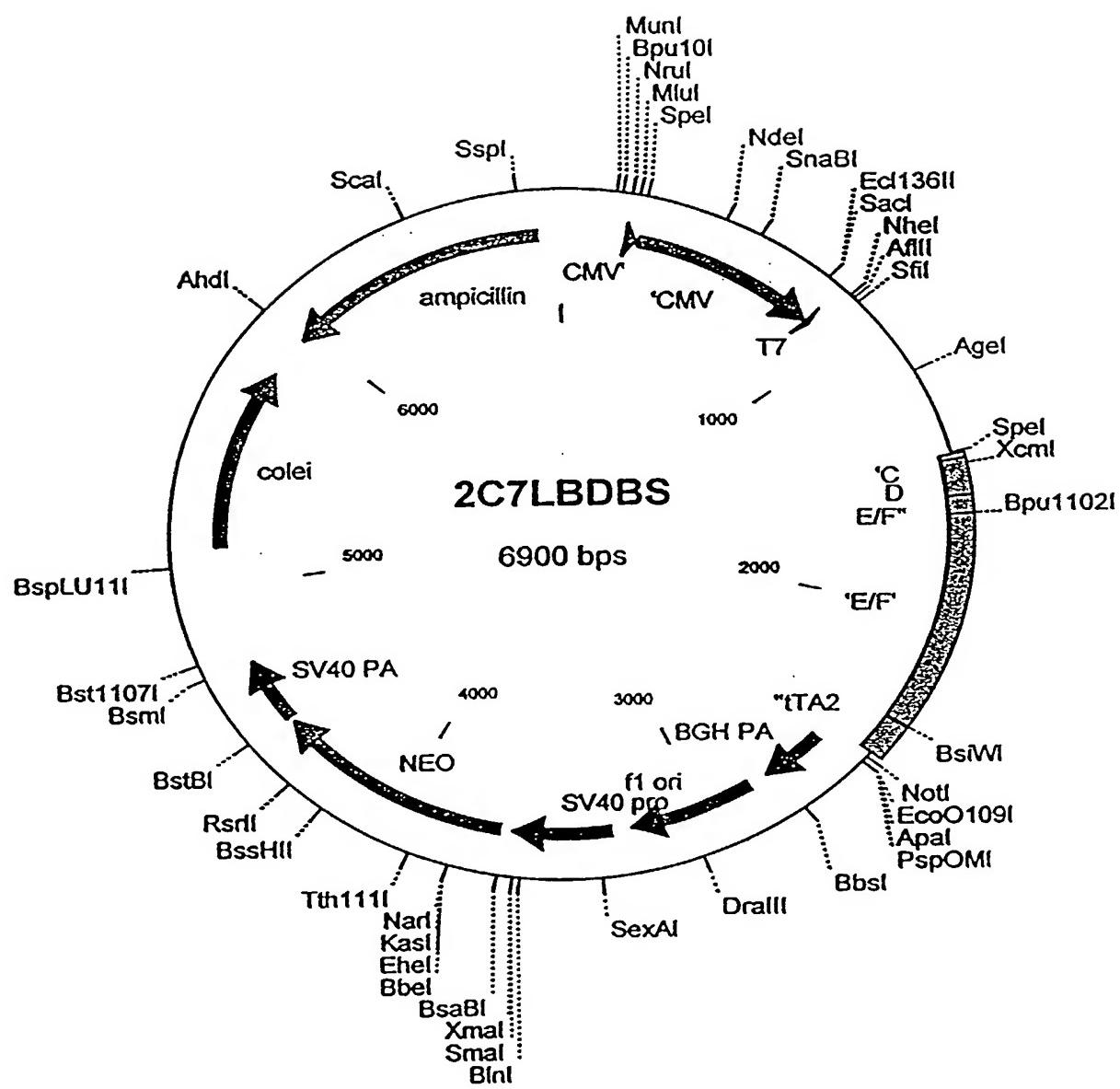


FIGURE 17

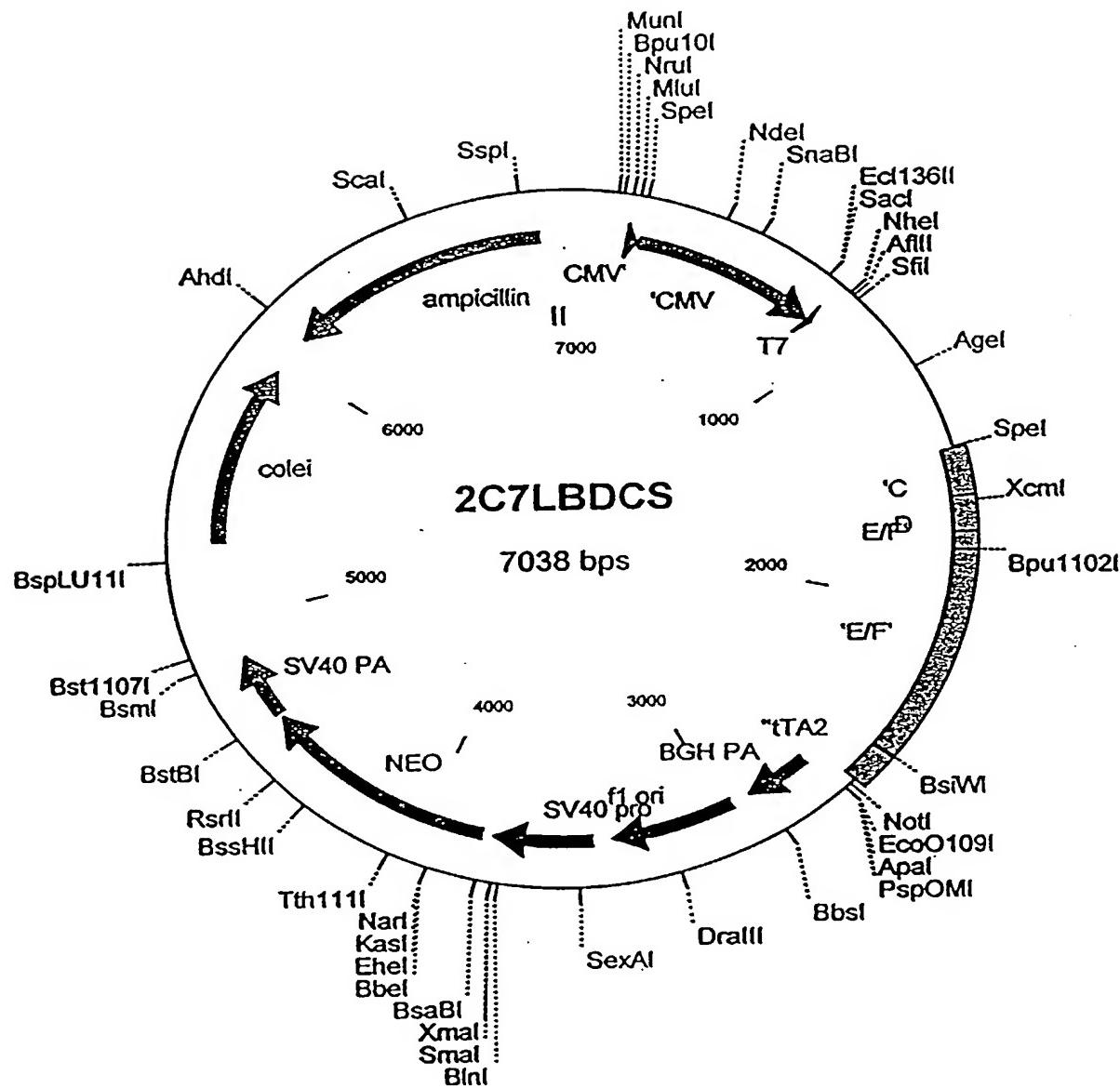


FIGURE 18

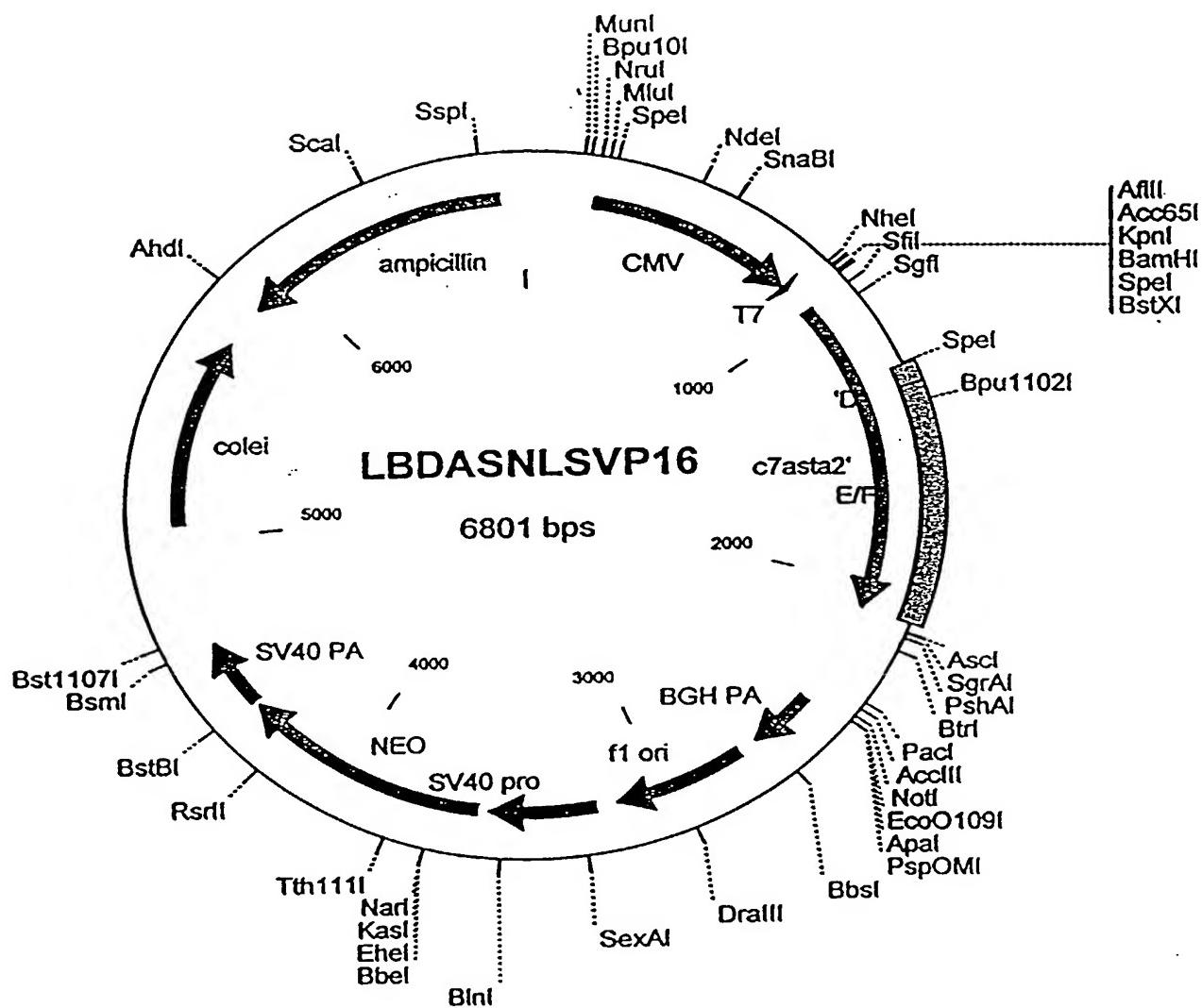


FIGURE 19

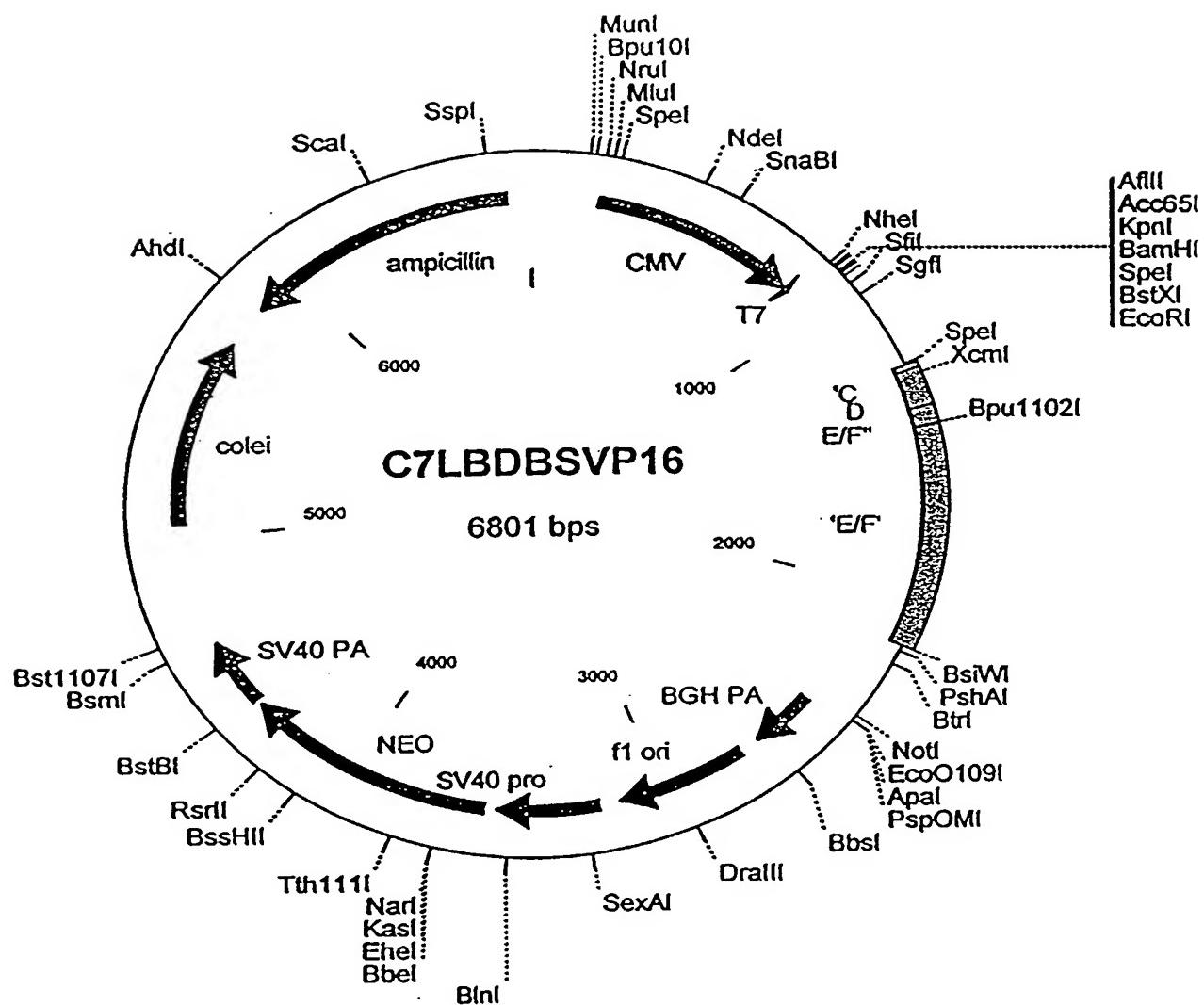


FIGURE 20

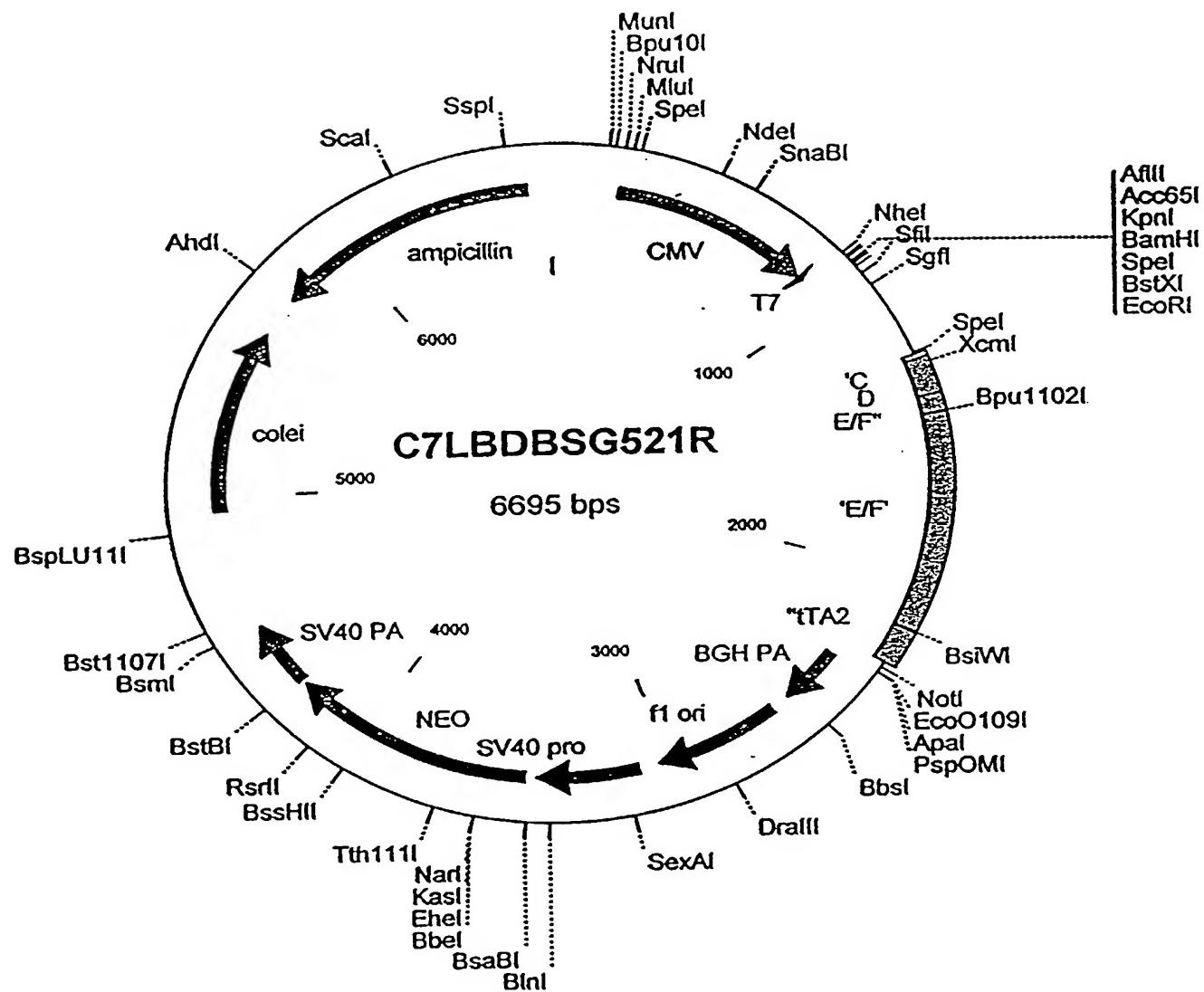


FIGURE 21

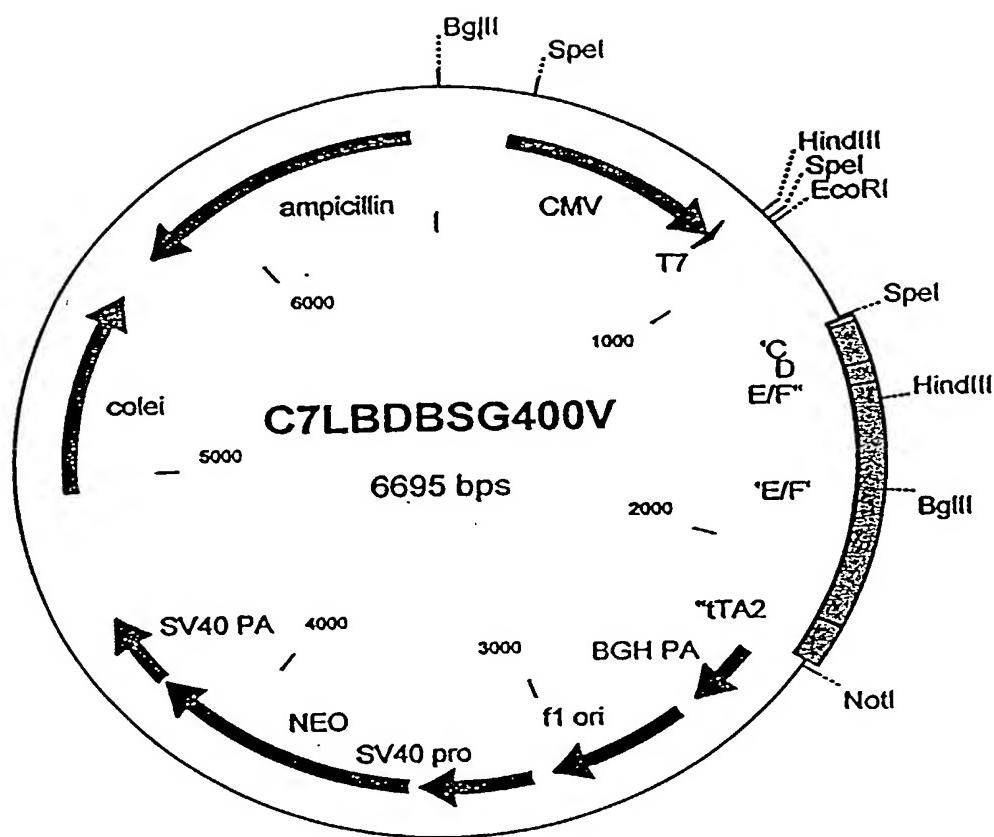
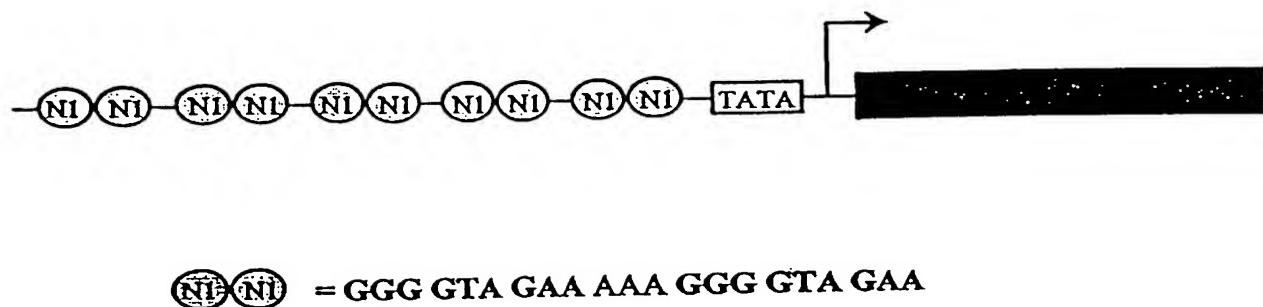
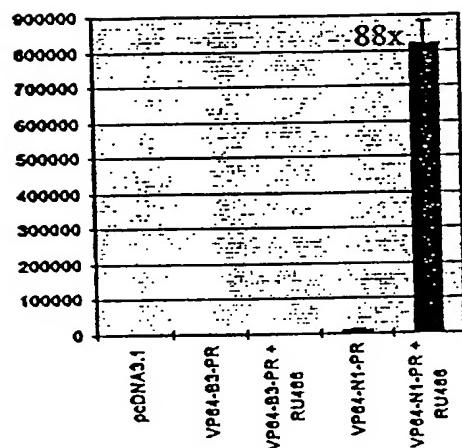


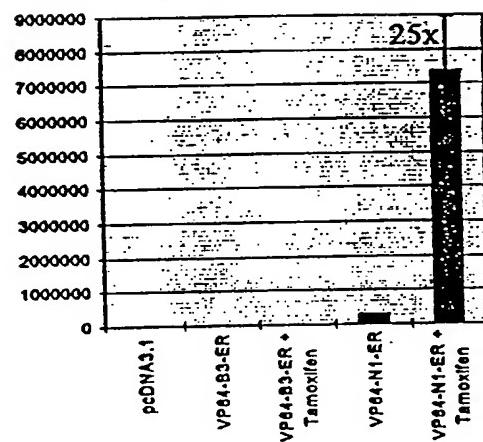
FIGURE 22

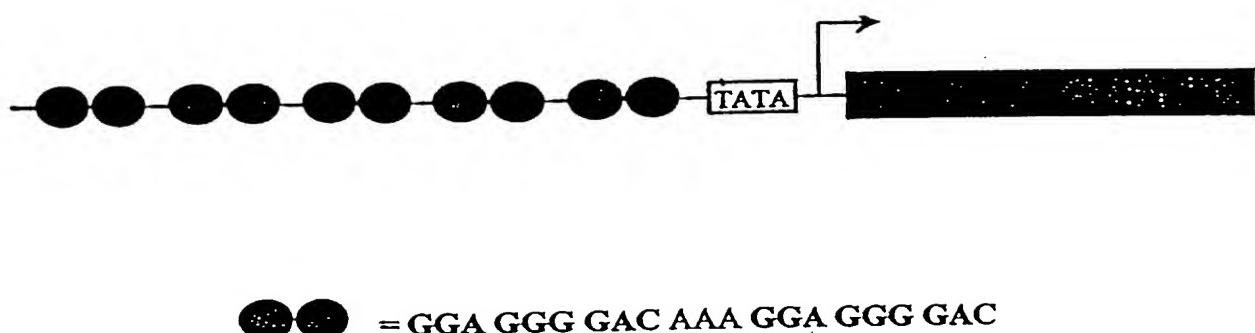
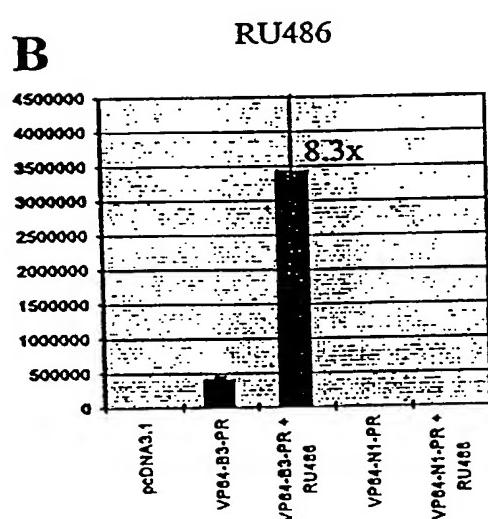
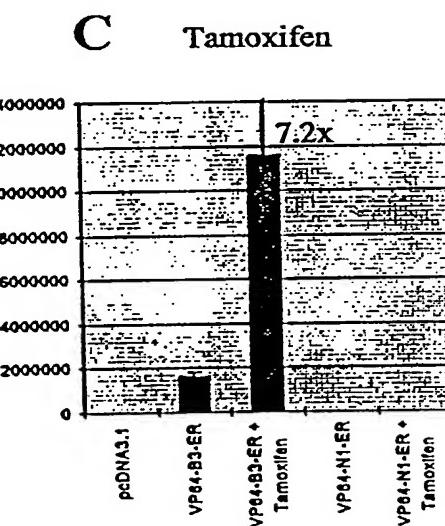
A**B**

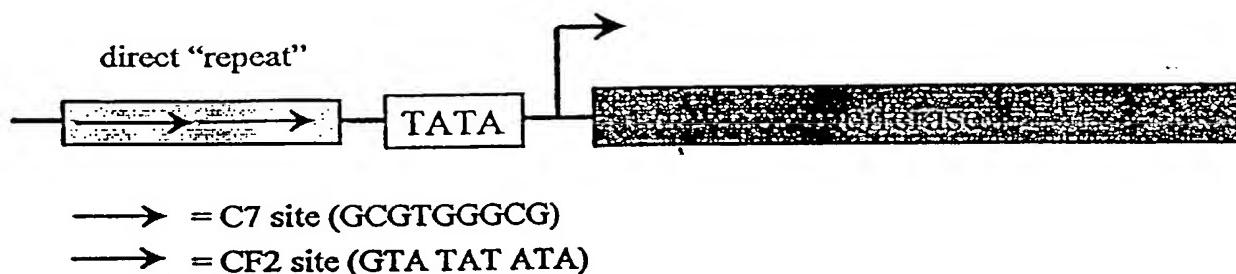
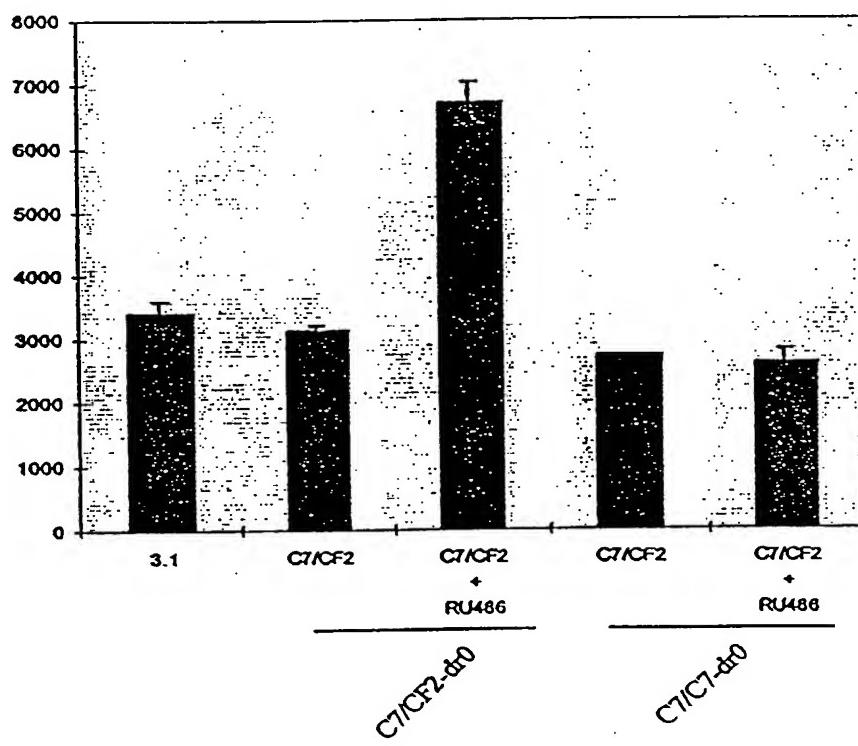
RU486

**C**

Tamoxifen

**FIGURE 23**

A**B****C****FIGURE 24**

A**B****FIGURE 25**

Left end shuttle plasmid for adenovirus vector construction

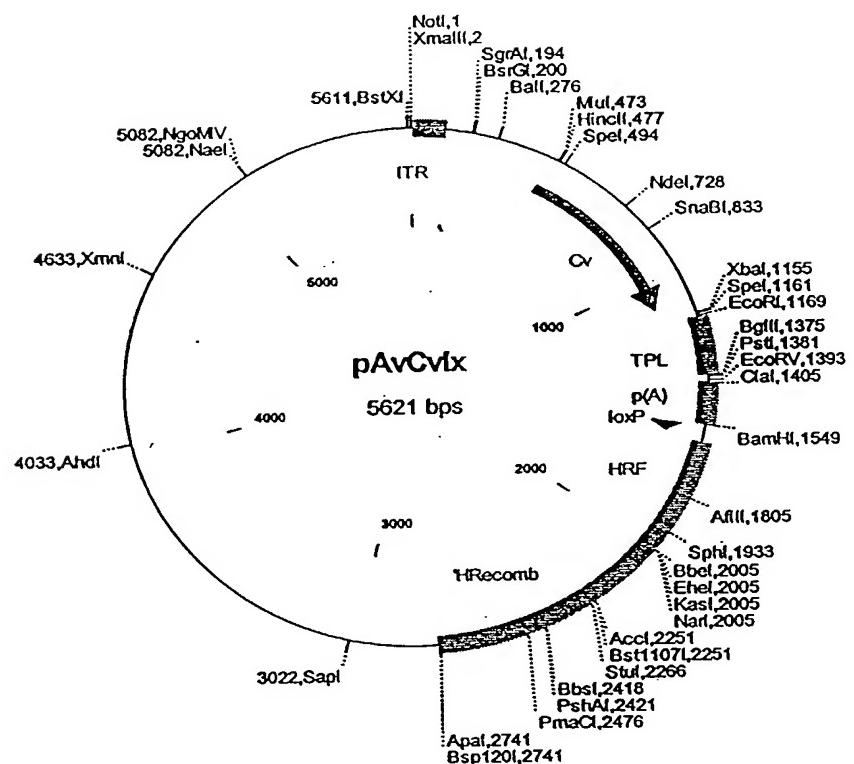
**FIGURE 26**

Figure 24. Right End Vector Genome-containing plasmid

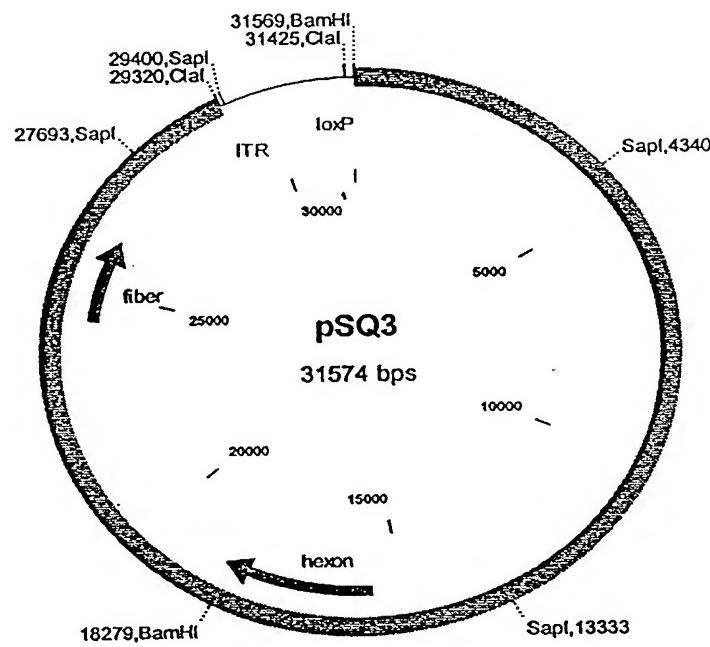


FIGURE 27

SEQUENCE LISTING

<110> Barbas, Carlos F., III
Kadan, Michael
Beerli, Roger

<120> LIGAND ACTIVATED TRANSCRIPTIONAL REGULATOR PROTEINS

<130> 22908-1227B

<140> Unknown

<141> 2000-06-02

<150> 09/433,042

<151> 1999-10-25

<160> 92

<170> PatentIn Ver. 2.0

<210> 1

<211> 6828

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Construct
2C7LBDAS

<400> 1

gacggatcgg gagatctccc gatcccstat ggtcgactct cagtacaatc tgctctgatg 60
ccgcatagtt aagccagtat ctgctccctg ctttgtgttt ggaggtcgct gagtagtgcg 120
cgagcaaat ttaagctaca acaaggcaag gcttgaccga caattgcatt aagaatctgc 180
ttagggttag gcgtttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt 240
gattattgac tagtattaa tagtaatcaa ttacggggtc attagttcat agcccatata 300
tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360
cccgccatt gacgtcaata atgacgtatg ttccatagt aacgccaata gggactttcc 420
attgacgtca atgggtggac tatttacggt aaactgccc cttggcagta catcaagtgt 480
atcatatgcc aagtacgccc octattgacg tcaatgacgg taaatggccc gcctggcatt 540
atgcccagta catgacccta tgggactttc ctacttggca gtacatctac gtattagtca 600
tcgcttattac catggatg cggttttggc agtacatcaa tggcgtgga tagcggttt 660
actcacgggg atttcaagt ctccacccca ttgacgtcaa tggaggttg tttggcacc 720
aaaatcaacg ggacttcca aatatgtcgta acaactccgc cccattgacg caaatggcg 780
gtaggcgtgt acgggtggag gtctatataa gcagagctct ctggctaact agagaaccga 840
ctgcttactg gcttatcgaa attaatacga ctcactatag ggagacccaa gctggctagc 900
gtttaaactt aagcttagat ctatggccca ggcggccctc gagccctatg cttgcccgt 960
cgagtccctgc gatccggctt tttctaagtc ggctgatctg aagcgccata tccgcattca 1020
cacaggccag aagcccttcc agtgtcgaat atgcattcggt aacttcagtc gtatgtacca 1080
ccttaccacc cacatccgca cccacacagg cgagaaggct tttgcctgtg acatttgg 1140
gagaaagttt gccaggagtg atgaacgc当地 gaggcatacc aaaatccata ccggtgagaa 1200
gccctatgtct tgccctgtcg agtcctgcga tcggccgtt tctaagtccgg ctgatctgaa 1260
gccccatccatc cgcattccaca caggccagaa gcccctccag tgcgaatat gcatgcgtaa 1320
cttcagtcgt agtgaccacc ttaccaccca catccgcacc cacacaggcg agaaggcctt 1380
tgccctgtgac atttgggaa ggaagtttc cagagtgat gaacgc当地 ggcataccaa 1440
aatccattta agacagaggg actctagaac tagttctgtct ggagacatga gagctgccaa 1500
ccttggcca agcccgctca tgcataaacg ctctaaaggaa aacagccctgg ccttgcct 1560
gacggccgac cagatggctca gtgccttgg ggtgctgag ccccccatac tctattccga 1620
gtatgtatcc accagaccct tcagtgaaat ttcgtatgtat ggcttactga ccaacctggc 1680
agacagggag ctggtcaca tgcataactg ggcaagagggtt gtcggaggct ttgtggattt 1740
gaccctccat gatcagggtcc accttctaga atgtgcctgg ctagagatcc tgatgatttt 1800
tctcgtctgg cgctccatgg agcaccaggaa gaagctactg tttgcctcta acttgctttt 1860
ggacaggaac caggaaaat gtgtagaggg catggtggag atcttcgaca tgctgctggc 1920

tacatcatct cggttccgca ttagtgaatct gcagggagag gagtttgtgt gcctcaaattc 1980
 tattattttg ctaattctg gagggtacac attttctgtcc agcacccctga agtctctggaa 2040
 agagaaggac catatccacc gaggctctgga caagatcaca gacacttga tccacactgat 2100
 ggccaaggca ggcctgaccc tgcagcagca gcaccagcg 2160
 cctctccacatcaggcaca ttagtgcacaa aggcatggag ctggcccagc tcctcctcat 2160
 caagaacgtg gtgccttcct atgacctgt gctggagatg catctgtaca gcatgaagtg 2220
 tgcccccact agccgtacgc cggccgacgc cctggacgc 2280
 ggcgacgco ctggacact tcgacctgga catgtcccg 2340
 cgacctggac atgtcggcgg ggtactaataa taagcggccg 2400
 ttaaaccgc ttagtgcct cgactgtggc ttcttagttgc 2460
 ctccccctg cttccctgaa ccctggaaagg tgccactccc 2520
 tgaggaaatt gcatcgatt gtctgagtag gtgtcattct 2580
 gcaggacago aagggggaggg attgggaaaga caatagcagg 2640
 ctctatggct tctgaggcgg aaagaaccag ctgggctct 2700
 ctgtagcggc gcattaagcg cggcgggtgt ggtgggtacg 2760
 tgccagcgcc cttagccccg ctcccttcgc tttctccct 2820
 cggttccctt cgtcaagtc taaaatcggg catccctta 2880
 acggcaccc gaccaaaaaa acttgatata ggggtatgg 2940
 ctgatagacg gttttcgcc ctttgacgtt ggagtcacg 2940
 gttccaaact ggaacaacac tcaaccctat ctcgtctat 3000
 tttgggatt tcggctatt gttaaaaaa tgagctgatt 3060
 ttaattctgt ggaatgtgt ttagtttaggg tggaaagt 3120
 cagaagatag caaagcatgc atctcaattt gtcagcaacc 3180
 ctccccagca ggcagaagta tgcaaaagcat gcatctcaat 3240
 gcccctaact ccccccattcc cggcccttaac tccggccagt 3240
 tggtctacta attttttttta ttatgcaga gggcggggcc 3300
 ccagaatgg tagggggctt tttttggggg ctttaggtt 3360
 ttgtatatcc attttcggat ctgatcaaga gacaggatga 3420
 acaagatgga ttgcacgcg gttctccggc cgctgggtg 3480
 ctgggcacaa cagacaatcg gctgtctga tgccggcgtg 3540
 gcccgggtt cttttgcgaa agaccgaccc gtcgggtgcc 3600
 ggcagcgcgg ctatcggtc tggccacgac ggggttcct 3660
 tgtactgaa gggggaaagg actggctgtt attggggcga 3720
 gtcatctcac ctgtccctg ccgagaaagt atccatcatg 3780
 gcatacgctt gatccggctt ctcgtccatt cgaccacca 3840
 agacgtact cggatggaa cgggtttttgt cgatcaggat 3900
 gggctcgcg ccagccgaac tggccggccag gctcaaggcg 3960
 tctctgtcg acccatggcg atgcctgtt gccaatata 4020
 ttctggattt atcgactgtg gccggctggg tggcgggac 4080
 ggctacccgtt gatattgtcg aagagcttgg cggcgaatgg 4140
 ttacggatcc gccgctcccg attcgacgcg catcgcttc 4200
 cttctgagcg ggactctggg gttcgaaatg accgaccaag 4260
 cgagatttcg attccaccgc cgccttctat gaaagggtgg 4320
 gacggccgtt ggtatgttcc ccaagcgggg gatctcatgc 4380
 aacctgtta ttgcagctta taatgttttca aaataaaagca 4440
 aataaaagcat ttttttact gcattctagt tgggtttgt 4500
 tatcatgtct gtataccgtc gacctctagc tagagcttgg 4560
 tttctgtgtt gaaattgtta tccgctcaca attccacaca 4620
 aagtgtaaag cctgggggtgc ctaatgagtg agctaactca 4680
 ctggccgctt tccagtcggg aaacctgtcg tgccagctgc 4740
 gcgggggagag gcggtttcg gcgagcggta tcagctact 4800
 tccacagaat caggggataa cgcaggaaag aacatgtgag 4860
 aggaacccgtt aaaaggccgc gttgtggcg ttttccata 4920
 catcacaaaaa atcgacgctc aagttagcagg tggcggaaacc 4980
 caggcgttcc cccctggaaag ctccctcggt cgctctccgt 5040
 ggataccgtt ccgccttctt cccttggga agcgtggcgc 5100
 aggtatctca gttcgggtgtt gtcgtttcgc tccaagctgg 5160
 gttcagcccg accgctgcgc ttatccgtt aactatcgtc 5220
 cacgacttat cgccactggc agcagccact ggttaacagga 5280
 ggcggtgcta cagagtctt gaaatggggg ccttaactacg 5340
 tttggtatctt ggcgtctgtt gaaagccgtt accttcggaa 5400
 tccggcaaac aaaccaccgc tggtagcggg gtttttttgc 5460
 cgcaaaaaaa aaggatctca aagagatctt ttgtatcttt 5520
 tggaaacggaa actcacgtt aaggattttt gtcgtttttttt 5580
 tagatccctt taaattaaaa atgaagttt aaatcaatctt 5640
 tggctgaca gttaccaatg ctaatcagt gaggcaccta 5700
 cgttcatcca tagttgcctg actccccgtc gtgttagataa 5760
 ttcacggcgat ctgtcttattt 5820
 ttcacggcgat ctgtcttattt 5880
 ctcacggcgat ggagggctta 5940

ccatctggcc ccagtgcgtc aatgataccg cgagacccac gctcacccgc tccagattt 6000
 tcagcaataa accaggccagc cggaaaggcc gaggcgagaa gtggctctgc aactttatcc 6060
 gcctccatcc agtctattaa ttgttgccgg gaagcttagag taagtagttc gccagttaa 6120
 agtttgcgca acgttgcgtc cattgctaca ggcatcggtg tgtcacgctc gtcgtttgg 6180
 atggcttcat tcagctccgg ttcccaacga tcaaggcgag ttacatgatc cccatgttg 6240
 tgcaaaaaaaag cggttagctc cttcggtctt ccgatcggtg tcagaagtaa gttggccgca 6300
 gtgttatcac tcatgtttaat ggcagcactg cataattctc ttactgtcat gccatccgta 6360
 agatgtttt ctgtgactgg tgtagtactca accaagtcat tctgagaata gtgtatgcgg 6420
 cgaccgagggt gtcgtcccc ggcgtcaata cggataata cccgcggcaca tagcagaact 6480
 ttaaaaatgc tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccc 6540
 ctgttgagat ccagttcgat gtaacccact cgtgcaccca actgatcttcc agcatcttt 6600
 actttcacca gcgttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga 6660
 ataaggcgca cacggaaatg ttgaataactc atactttcc ttttcaata ttattgaagc 6720
 atttacagg gttattgtct catgagcgga tacatattt aatgtattta gaaaaataaaa 6780
 caaatagggg ttccgcgcac atttcccgaa aaagtgcac ctgacgctc 6828

<210> 2
 <211> 6900
 <212> DNA
 <213> Artificial Sequence

<220>

<220>
 <223> Description of Artificial Sequence: Construct
 2C7LBDBS

<400> 2
 gacggatcg ggagatctccc gatccccat ggtcgactct cagtacaatc tgctctgatg 60
 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gагtagtgcg 120
 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatt aagaatctgc 180
 tttagggttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatacg cggtgacatt 240
 gattattgac tagttattaa tagtaatcaa ttacgggggtt attagttcat agcccatata 300
 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaaacgacc 360
 cccggccatt gacgtcaata atgacgtatg ttcccatagt aacgcataa gggactttcc 420
 attgacgtca atgggtggac tattttacggt aaactgccc cttggcagta catcaagtgt 480
 atcatatgccc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540
 atgcccagta catgaccccta tgggactttc ctacttggca gtacatctac gtattagtca 600
 tcgcttattac catggtgatg cgggtttggc agtacatcaa tggcgtgga tagcggtttg 660
 actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttgg ttttggcacc 720
 aaaatcaacgg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggg 780
 gttaggcgtgt acgggtggag gtctatataa gcaagactct ctggctacta agagaaccca 840
 ctgcttactg gtttacgaa attaatacgat ctcactatag ggagacccaa gctggcttagc 900
 gtttaaactt aagcttagat ctatggccca ggccggccctc gagccctatg cttggccctgt 960
 cgagtccctgc gatcggcgct tttctaagtc ggctgatctg aagcgcata tccgcatcca 1020
 cacaggccag aaggccccc agtgtcgaat atgcatcgat aacttcagtc gtatgtacca 1080
 ctttaccacc cacatccgca cccacacagg cgagaagcct tttgcctgtg acattttgtgg 1140
 gaggaagttt gccaggagtg atgaacgca gaggcatacc aaaatccata cccgtgagaa 1200
 gcccctatgct tggcctgtcg agtcctgcga tcggcgctt tctaagtctgg ctgatctgaa 1260
 gcgccatatac cgcattccaca caggccagaa gcccctccag tgcgaaat gcatgcgtaa 1320
 cttcagtcgt agtgaccacc ttaccacca catccgcacc cacacaggcg agaaggccctt 1380
 tggctgtgac attttgggaa ggaatgttgc caggagtatc gaacgcgaa ggcataccaa 1440
 aatccattta agacagaggg actcttagaac tagtggaccga agaggaggaa gaatgttggaa 1500
 acacaaggcg cagagagat atggggagggg cagggggtgaa gtggggcttg ctggagacat 1560
 gagagctgcc aaccttggc caagcccgct catgatcaaa cgctctaaga agaacagcct 1620
 ggccttgcctc ctgacggccg accagatgtt cagtgccctt ttggatgtctg agccccccat 1680
 actcttattcc gagtatgatc ctaccagacc cttcagtgaa gtttcgtatg tgggttact 1740
 gaccaacctg gcagacaggg agtcgggtca catgatcaac tggcgaaga ggggtccagg 1800
 ctttgcgttat ttgacccctcc atgatcagggt ccaccccttca gaatgtccct ggcttagat 1860
 cctgtatgtt ggtctcgctt ggcgtccat gggacccca gggaaatctac tgggtctcc 1920
 taacttgcgtc ttggacagga accaggaaaa atgtgttagag ggcatgttgg agatcttoga 1980
 catgctgtgtc gtcatacatc ctgggttccg catgatgaat ctgcaggagg aggagttgt 2040
 gtgcctcaaa tctatttattt tgcttaattt tgagggtgtac acatttctgt ccagcaccc 2100
 gaagtctctg gaagagaagg accatatcca ccgagtcctg gacaagatca cagacacttt 2160
 gatccacactg atggccaagg caggcctgac cctgcagcag cagcaccagg ggctggccca 2220
 gctcctcctc atcctctccc acatcaggca catgagtaac aaaggcatgg agcatctgta 2280
 cagcatgaag tgcaagaacg tggtgcacccct ctatgacgctc ctgctggaga 2340

ccaccgccta	catgcgccc	ctagccgtac	gccggccgac	gccctggacg	acttcgacct	2400
ggacatgctg	ccggccgacg	ccctggacga	cttcgacctg	gacatgctgc	cggccgacgc	2460
cctggacgac	ttcgacctgg	acatgctgcc	ggggtaacta	agtaagcggc	cgctcgagtc	2520
tagaggccc	gtttaaaccc	gctgatcagc	ctcgactgtg	ccttctagtt	gccagccatc	2580
tgttgttgc	ccctcccccg	tgccctccct	gaccctggaa	ggtgccactc	ccactgtcct	2640
ttccataaa	aatgaggaaa	ttgcatcga	ttgtctgagt	aggtgtcatt	ctattctggg	2700
gggtggggtg	gggcaggaca	gcaaggggga	ggattgggaa	gacaatacgca	ggcatgctgg	2760
ggatcggtg	ggctctatgg	cttctgaggc	ggaaaagaacc	agctggggct	ctagggggta	2820
tccccacgcg	ccctgttagc	gcmcattaag	cgcggcgggt	gtgggtgtta	cgcgcagcgt	2880
gaccgctaca	cttgcacgc	ccctagcgc	cgctccccc	gctttctcc	cttccttct	2940
cgccacgttc	gccggcttc	cccgtaagc	tctaaatcgg	ggcatccctt	tagggttccg	3000
atttagtgt	ttacggcacc	tcgaccccaa	aaaacttgat	tagggtgatg	gttcacgtag	3060
tggccatcg	ccctgtataga	cgtttttcg	cccttgcacg	ttggagtcca	cgttctttaa	3120
tagtggactc	ttgttccaaa	cttgaacaac	actcaaccct	atctcggtct	attcttttga	3180
tttataaggg	attttgggg	tttcggccta	ttggttaaaa	aatgagctga	tttaacaaaaa	3240
atthaacgcg	aattaattct	gttgaatgtg	tgtcagttag	ggtgtggaaa	gtccccaggc	3300
tccccaggca	ggcagaagta	tgc当地atcgat	gcatctcaat	tagtcagca	ccaggtgtgg	3360
aaagtcccc	ggctcccccag	caggcagaag	tatgc当地aa	atgcatctca	attagtcagc	3420
aaccatagtc	ccgcccctaa	ctccgccccat	cccgccccct	actccgccc	gttccgccc	3480
ttctccgccc	catggctgac	taattttttt	tatttatgca	gaggccgagg	ccgcctctgc	3540
ctctgagcta	ttccagaagt	agtgaggagg	cttttttgg	ggcctaggct	tttgc当地aaa	3600
gctcccgaaa	gcttgatata	ccattttcgg	atctgatcaa	gagacaggat	gaggatcggt	3660
tcgcatgatt	gaacaagatg	gattgcacgc	aggttctccg	gccgcttggg	tggagaggct	3720
attcggttat	gactgggcac	aacagacaaat	ccgctgtct	gatggccccc	tgttccggct	3780
gtcagcgcag	gggcgcgg	ttctttttgt	caagaccgc	ctgtccgggt	ccctgaatga	3840
actgcaggac	gaggcagcgc	ggatatcgt	gctggccacg	acgggcgttc	cttgcgcagc	3900
tgtgtcgac	gttgc当地tgc	aagcgggaa	ggactggctg	ctattggcg	aagtgc当地gg	3960
gcaggatctc	ctgtcatctc	accttgc	tgccgagaaa	gtatccatca	tggctgatgc	4020
aatgcggcgg	ctgcatacgc	ttgatccggc	tacctgccc	tgc当地acc	aagc当地aaa	4080
tcgcatcgag	cgagcacgt	ctcggatgg	agccggctt	gtc当地tgcagg	atgatctgg	4140
cgaagagcat	cagggctcg	cgccagccga	actgttcgccc	aggctcaagg	cgc当地atgc	4200
cgacggcgg	gatctcgctg	tgacccatgg	cgatgcctgc	ttgccc当地ata	tcatggtgg	4260
aaatggccgc	ttttctggat	tcatgcactg	tggccggctg	ggttgtggcgg	accgctatca	4320
ggacatagcg	ttggctaccc	gtgatattgc	tgaagagctt	ggcggcgaat	gggtgaccc	4380
cttcctcggt	ctttacgta	tcgcccgtcc	cgattcgcag	cgcatcgct	tctatcgct	4440
tctttagcag	ttcttctgag	cgggactctg	gggttc当地aa	tgaccgacca	agc当地acgccc	4500
aacctgccc	cacgagattt	cgattccacc	ccgc当地ttct	atgaaagg	gggcttc当地gg	4560
atcgttttcc	gggacggcgg	ctggatgatc	ctccagcgc	gggatctcat	gctggagttc	4620
ttcgcccacc	ccaacttgtt	tattgcagct	tataatgg	acaataaa	caatagc	4680
acaatattca	caaataaagc	attttttca	ctgc当地tctca	gttgc当地tgc	gtccaaactc	4740
atcaatgtat	cttacatgt	ctgtataccg	tcgaccc	gctagagctt	ggc当地atca	4800
tggctatgc	tgttccctgt	gtgaaatttg	tatccgct	caattccaca	caacatacg	4860
ggccggaa	taaagtgtaa	aggctgggg	gccttaatgag	ttagctact	cacattaatt	4920
gcgttgcgt	cactgcccgc	tttccagtcg	ggaaacac	cgtgccc当地ct	gcattaatga	4980
atcgccaa	gcgc当地gg	aggcggg	cggc当地gg	tatcagctca	ctcaaaggcg	5040
gtaatacggt	tatccacaga	atcaggggat	aacgc当地ggaa	agaacatgt	agcaaaaaggc	5100
cagcaaaaagg	ccaggaaccg	taaaaaggcc	gcgttgc	cg	tttcc	5160
ccccctgacg	agcatcacaa	aaatcgacgc	tcaagtc	gttgc当地gg	ccgc当地agg	5220
ctataaagat	accaggcg	tccccc	agctccctcg	tcgct	tgttccgacc	5280
ctggcgtta	ccggatacc	gtccgc	ctccctt	gaagcgttgc	gttctc当地aa	5340
tgctcacgct	gttaggtatc	cgttgc	ttgtc	gttcc	gggctgtgt	5400
cacgaacccc	ccgttca	cgaccgc	gcctt	gttactatcg	tcttgc当地tcc	5460
aacccggtaa	gacacgactt	atgc当地act	gcagc	ctggtaac	gattagc	5520
gcfaggtatg	taggc当地tgc	tacagagtt	ttga	ggc当地acta	cggtc当地act	5580
agaaggacag	tatttggat	ctgc当地tctg	ctga	ttac	aaaaagg	5640
ggtagctt	gatccggca	acaaaccacc	gctgg	ttttt	tgttgc当地ag	5700
cagcagatta	cgcgca	aaaaggatct	caaga	tcttgc	tttacgggg	5760
tctgacgctc	agtgg	aaactcac	taagg	tgttgc	attatcaaaa	5820
aggatctca	cctagatc	tttaaattaa	aaat	ttaa	ctaaaggata	5880
tatgagtaaa	cttgc当地tgc	cgttaccaa	tgctt	gtgagg	gttctc	5940
atctgtctat	ttcgttcatc	catagttgc	tgact	tcgt	actac	6000
cgggagggct	taccatctgg	cccc	gcaat	cgcg	gagac	6060
gctccagatt	tatcagcaat	aaacc	gggg	ccgagc	gaggtgtt	6120
gcaactttat	ccgc当地ccat	ccagtctt	aatttgc	ggaa	agtagt	6180
tcgcccagtt	atagttgc	caacgtt	gccatt	cagg	catcg	6240
tctgtcg	gtatggctt	attcag	ggtt	gatca	aggcg	6300
tcccccattgt	tgtgcaaaaa	agcggtt	ac	ctccgat	tgtcaga	6360

aagtggccg cagtgttatac actcatgttt atggcagcac tgcataattc tcttactgtc 6420
atgccatccg taagatgctt ttctgtact ggtgagact caaccaagtc attctgagaa 6480
tagtgtatgc ggcgaccgag ttgctctgc ccggcgtaa tacgggataa taccgcgcca 6540
catagcagaa cttaaaaagt gctcatcatt ggaaaacgtt cttcgccccg aaaactctca 6600
agatcttac cgctgttgag atccagttcg atgttaaccca ctcgtgcacc caactgatct 6660
tcagcatctt ttactttcac cagcgtttct gggtgagcaa aaacaggaag gaaaaatgcc 6720
gcaaaaaagg gaataaggc gacacggaaa tggtaatac tcatactttt ccttttcaa 6780
tattattgaa gcatttatca gggttatgt ctcatgagcg gatacatatt tgaatgtatt 6840
tagaaaaata aacaatagg ggttccgcg acatttcccc gaaaagtgcc acctgacgtc 6900

<210> 3
<211> 7038
<212> DNA
<213> Artificial Sequence

<220>

<220>
<223> Description of Artificial Sequence: Construct
2C7LBDCS

<400> 3
gacggatcgg gagatctccc gatcccattt ggtcgactct cagtacaatc tgctctgatg 60
ccgcatagtt aagccagttt ctgctccctg ctttgtgttt ggaggctcgct gaggatgtcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatt aagaatctgc 180
ttagggttag gcggtttcgct ctgcgttcgc atgtacgggc cagatatacg cggtgacatt 240
gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata 300
tgaggttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaaacgacc 360
cccgccattt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420
attgacgtca atgggtggac tatttacggt aaactgccc cttggcagta catcaagtgt 480
atcatatgccc aagtacgccc cctattgacg tcaatgacgg taaaatggccc gcctggcatt 540
atgcccagta catgaccccta tgggactttc ctacttggc gtacatctac gtattagtc 600
tcgcttattac catggtgatg cgggttttgc agtacatcaa tgggcgttga tagcggttt 660
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttgc ttttggcacc 720
aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggcg 780
gtaggcgtgt acgggtggag gtctatataa gcagagctct ctggcttaact agagaaccca 840
ctgcttactg gcttatcgaa attaatacga ctcaactatag ggagacccaa gctggctagc 900
gtttaaactt aagcttagat ctatggccca ggccgcctc gagccctatg ctggccctgt 960
cgagtcctgc gatgcgcgtc ttcttaagtc ggctgatctg aagcgcata tccgcatacca 1020
cacaggccag aaggctttcc agtgtcgaat atgcattgcgt aacttcagtc ttagtgcacca 1080
ccttaccacc cacatccgca cccacacagg cgagaaggct tttgcgttg acatttgggg 1140
gaggaagttt gccaggagtg atgaacgcaaa gggcataacc aaaatccata ccggtgagaa 1200
gcctatgtct tgcctgtcg agtccctgcga tcgcccctt tctaagtgcg ctgtatctgaa 1260
gcccattatc cgcattccca caggccggaa gcccattccag tgcataatgcgatgcgtaa 1320
cttcagtcgt agtgaccacc ttaccacccaa catccgcacc cacacaggcg agaaggcctt 1380
tgcctgtgac atttgtggaa ggaagtttgc caggagtgtat gaacgcacca ggcataccaa 1440
aatccattta agacagaggg actctagaac tagtagtatt caaggacata acgactat 1500
gtgtccagcc accaaccagt gcaccatga taaaaacagg aggaagagct gccaggccctg 1560
ccggctccgc aaatgtacg aagtggaaat gatgaaaggat gggatacgcg aagaccgaaag 1620
aggagggaga atgttgcac acaagcgcac gagagatgtat gggggggca ggggtgaagt 1680
ggggctgtctt ggagacatga gagctgcac ccttggcca agcccgtca tgatcaaacg 1740
ctctaagaag aacagcctgg ctttgtccct gacggccgcac cagatgttc tgccctgtt 1800
ggatgtcgag cccccccatc tcttattccga gtatgtatcctt accagaccct tcagtgaagc 1860
ttcgatgtatc ggcttactgcca acaacccgtt agacaggggat ctggttcaca tgatcaactg 1920
ggcgaagagg gtgcaggctt ttgtggattt gaccctccat gatcagggtcc accttctaga 1980
atgtgcctgg ctagagatcc tgatgattgg tctcgctctgg cgctccatgg agcaccggc 2040
gaagctactg ttgtctctt acttgctttt ggacagggaaat cagggaaaat gtgttagaggg 2100
catggtgaaat atcttcgaca tgctgtgc tacatcatct cggttccgc tgatgaatct 2160
gcaggggagag gagttgtgt gcctcaatc tattatggat cttaaatttcg gagtgacac 2220
atttctgtcc agcaccctgaa agtctctgg agagaaggac catatccacc gagtcctgg 2280
caagatcaca gacacccatc tccacctgtat ggccaaaggca ggcctgcacc tgcagcagca 2340
gcaccaggcg ctggcccgatc ttctccctat cctctcccaatc atcaggcaca tgagtaacaa 2400
aggcatggag catctgtaca gcatgaagtg caagaacgtg gtgcctctt atgacctgtc 2460
gtggagatgt ctggacgccc accgcctaca tgcccccact agccgtacgc cggccgacgc 2520
cctggacgac ttcgacctgg acatgtctgc ggccgacgccc ctggacgact tcgacactgga 2580
catgtctccg gccgacgccc tggacgactt cgacccgtt atgctgcgg ggtactaag 2640
taagcggccg ctcgagtcata gaggcccgatc ttaaaccgcg tgatcagccct cgactgtgcc 2700

tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat gtaacccact 6780
 cgtgcaccca actgatctc agcatcttt actttcacca gcgttctgg gtgagcaaaa 6840
 acaggaaggc aaaatgccgc aaaaaagggc ataagggcga cacggaaatg ttgaatactc 6900
 atactcttc ttttcaata ttattgaagc atttatcagg gtattgtct catgagcgga 6960
 tacatatttgc aatgtattta gaaaaataaa caaatagggg ttccgcgcac atttccccga 7020
 aaagtgcac ctgacgtc 7038

<210> 4
 <211> 1496
 <212> DNA
 <213> Artificial Sequence

<220>

<220>
 <223> Description of Artificial Sequence: Construct
 C7PBDVP16

<400> 4

ggtacccggat	ccgcacccat	ggcccaggcg	gccctcgagc	cctatgcttg	ccctgtcgag	60
tccctgcgtc	ggcgcttttc	taagtcggct	gatctgaagc	gccatatccg	catccacaca	120
ggccagaagc	ccttcagtg	tcgaatatgc	atgcgttaact	tcagtcgttag	tgaccacctt	180
accaccccaca	tccgcaccca	cacaggcgag	aagccttttg	cctgtgacat	ttgtgggagg	240
aagtttgcca	ggagtgatga	acgcaaggagg	cataccaaaa	tccatattaag	acagaaggac	300
tcttagaacta	gtggccaggc	cggccgcgtc	gaccagaaaa	agttcaataa	agttaggtt	360
gtgagagcac	tggatgtgt	tgtctccca	cagccagtgg	gcgttccaaa	tgaaagccaa	420
gccttaaggcc	agagattcac	tttttcacca	ggtcaagaca	tacagttgat	tccaccactg	480
atcaacctgt	taatgagcat	tgaaccagat	gtgatctatg	caggacatga	caacacaaaa	540
cctgacaccc	ccagttcttt	gctgacaagt	cttaatcaac	taggcggag	gcaacttctt	600
tcagtagtca	agtggctcaa	atcattgcca	ggttttcgaa	acttacatat	tgatgaccag	660
ataactctca	ttcagttttc	ttggatgagc	ttaatgggt	ttggctctagg	atggagatcc	720
tacaaaacacg	tcagtgccaa	gatgctgtat	tttgcacctg	atctaatact	aatgaacag	780
cggatgaaag	aatcatcatt	ctattcatta	tgcccttacca	tgtggcgat	cccacaggag	840
tttgtcaagc	ttcaagtttag	ccaagaagag	ttcctctgtt	tgaaagtatt	gttacttctt	900
aatacaattc	cttttggaaagg	gctacgaatg	caaaccagg	ttgaggagat	gaggtcaagc	960
tacatttagag	agctcatcaa	ggcaatttgg	ttgaggcaaa	aaggagttgt	gtcgagctca	1020
cagcgtttct	atcaacttac	aaaacttctt	gataacttgc	atgatcttg	caaacaactt	1080
catctgtact	gcttgaatac	atttatccag	tcccggcac	tgagtgttga	atttccagaa	1140
atgatgtctg	aagttattgc	tgggtcgacg	gctagcccga	aaaagaaacg	caaagttgg	1200
cgcgcggcgc	ctccccccac	cgatgtcagc	ctgggggacg	agtcactt	agacggcgag	1260
gacgtggcga	tggcgcatgc	cgacgcgcta	gacgatttcg	atctggacat	gttgggggac	1320
ggggatttccc	cggttccggg	atttaccccc	cacgactccg	ccccctacgg	cgctctggat	1380
atggccgact	tgcgagtttgc	gcagatgttt	accgatgccc	ttggatttgc	cgagtagcgt	1440
ttaattaact	acccgtacga	cgttccggac	tacgcttctt	gagaattcgc	ggccgc	1496

<210> 5
 <211> 6746
 <212> DNA
 <213> Artificial Sequence

<220>

<220>
 <223> Description of Artificial Sequence: Construct
 C7LBDAL

<400> 5

gacggatcg	gagatctccc	gatccctat	ggtcgactct	cagtacaatc	tgctctgtat	60
ccgcatagtt	aagccagtt	ctgctccctg	cttgcgtgtt	ggaggtcgct	gagtagtgcg	120
cgagcaaaat	ttaagctaca	acaaggcaag	gcttgaccga	caatttgcatt	aagaatctgc	180
tttagggttag	gcgttttgcg	ctgcttcgcg	atgtacgggc	cagatatacg	cggtgacatt	240
gattatttgc	tagttat	tagtaatcaa	ttacggggtc	attagttcat	agccatata	300
tgaggttccg	cgttacataa	cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	360
cccgccccatt	gacgtcaata	atgacgtatg	ttccctatgt	aacgccaata	gggactttcc	420
attgacgtca	atgggtggac	tatattacgt	aaactgccc	cttggcagta	catcaagtgt	480
atcatatgccc	aagtacgccc	cctattgacg	tcaatgacgg	taatggccc	gcctggcatt	540
atgcccagta	catgacctta	tgggactttc	ctacttggca	gtacatctac	gtattagtc	600

tcgctattac catggtgatg cggtttggc agtacatcaa tggcggtgga tagcggttg 660
 actcacgggg atttccaagt ctccacccca ttgacgtcaa tggaggttg tttggcacc 720
 aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggcg 780
 gtaggcgtgt acggtgggag gtctataaa gcagagctct ctggctaact agagaaccca 840
 ctgcttactg gcttatcgaa attaatacga ctcactatag ggagacccaa gctggctagg 900
 gttaaactt aagcttggta ccagagctcg atccactagt ccagtgttgt ggaattcctg 960
 cagccccggg gatctatggc ccaggcgccc ctcgagccct atgcttgcct tgcgagtcc 1020
 tgcgatcgcc gctttctaa gtcggctgt ctgaagcgcc atatccgcat ccacacaggc 1080
 cagaaggcctt tccagtgta aatatgcata cgtaacttca gtcgtagtga ccaccttacc 1140
 accacatcc gcacccacac aggcgagaag cctttgcct gtgacattt tgggaggaag 1200
 tttgccagga gtatgaacg caagaggcat accaaaatcc attaagaca gaggactct 1260
 agaacttagt ctgctggaga catgagagct gccaacctt ggccaagccc gctcatgatc 1320
 aaacgctcta agaagaacag cctggcttg tccctgacgg ccgaccagat ggtcagtgcc 1380
 ttgttggatg ctgagccccc catactctat tccgagatg atcctaccag acccttca 1440
 gaagcttcga tgatggcctt actgaccaac ctggcagaca gggagcttgt tcacatgatc 1500
 aactgggcga agagggtgcc aggcttgcg gattgaccc tccatgatca ggtccaccc 1560
 ctgaatgtg cctggctaga gatcctgtat attgtctcg tctggcgct catggagcac 1620
 ccagggaaagc gactgttgc tccataacttgc ctctggaca ggaaccaggg aaaatgtgt 1680
 gagggcatgg tggagatctt cgacatgtcg ctggctacat catctcggtt ccgcattat 1740
 aatctgcagg gagaggagtt tgtgtgcctt aaatctatta tttgcttaa ttctggagtg 1800
 tacacatttc tgtccagcac cctgaagttt ctggaaagaga aggaccatat ccacccgagtc 1860
 ctggacaaga tcacagacac ttgtatccac ctgtatggcca aggaggcctt gaccctgcag 1920
 cagcagcacc agcggctggc ccagctccct ctcatccctt cccacatcag gcacatgagt 1980
 aacaaaaggca tggagcatct gtacagcatg aagtgcaga aactggatcc cctctatgac 2040
 ctgctgctgg agatgctgga cgcccacccgc ctacatgcgc gctctactt atcgcatcc 2100
 tccgtggagg agacggacca aagccacttgc gccactgcgg ctggacacttcc 2160
 ttgaaaaagt attacatcac gggggagggc gagggtttcc ctggacacttcc 2220
 gccgacgccc tggacgactt cgacccgttgc atgctggccg cgcacgccc 2280
 gacctggaca tgctggccgc cgacgcccctg gacgacttc acctggacat gctgcccggg 2340
 taactaagta agcggccgtt cgagttctaga gggcccggtt aaacccgctg atcagcctcg 2400
 actgtgcctt ctatggcca gccatctttt gtttgccttcc ccccggtgcc 2460
 ctggaaagggtt ccactccac tgcctttcc taataaaatg aggaatttgc atcgcattgt 2520
 ctgagtaggtt gtcattctat tctgggggtt ggggtggggc aggacagcaa gggggaggat 2580
 tggaaagaca atagcaggca tgctggggat ggggtggggc ctatggcttc tgaggcggaa 2640
 agaaccagct ggggtcttag ggggtatccc caccgccttgc gtagcggcgc attaagcgcg 2700
 ggggtgtgg tggtaacgcg cagcgttgc gctacacttgc ctagccccc 2760
 ctttcgttcc tttcccttc ctttcgcgc acgttgcggc gcttccccc tcaagctcta 2820
 aatcggggca tccctttagg gttccgattt agtgcatttac ggcacctcga ccccaaaaaa 2880
 cttgatttagg gtatggttc acgttagtggg ccatgccttgc gatagacggt tttgcccct 2940
 ttgacgttgg agtccacgtt cttaataatg ggactcttgc tccaaactgg aacaacactc 3000
 aaccctatct cggcttattt ttttgcattt taagggattt tggggatttc ggcctattgg 3060
 ttaaaaaatg agtgcattt aaaaaatcc aacgcgaattt aattctgtgg aatgtgtgtc 3120
 agttaggggtg tggaaagtcc ccaggctcc caggcaggca gaatgtgca aagcatgcat 3180
 ctcaattagt cagcaaccag gtgtggaaag tcccccaggctt cccagcagg cagaagtatg 3240
 caaaacatgc atctcaatttgc gtcagcaaccatcccttcc ccttaactcc gcccatcccc 3300
 cccctaactc cggccagttc cggccatttcc cgcgccttgc gctgactaat ttttttattt 3360
 tatgcagagg cccaggccgc ctctgccttgc gagtatttcc agaagtagtg aggaggctt 3420
 tttggaggcc taggcttttgc caaaaagctc ccgggagctt gtatatccat ttctggatct 3480
 gatcaagaga caggatgagg atcggttgcg atgattgaac aagatggatt gcacgcaggt 3540
 tctccggccg cttgggtggc gaggctatttgc ggctatgact gggcacaaca gacaatcgcc 3600
 tgctctgtat cggccgttgc cggcgtgtca ggcgcaggggc gcccgggttct tttgtcaag 3660
 accacatgtt cgggtggccctt gaatgaaactt caggacggggc cagcgcggct atcgtggctg 3720
 gccacgacgg gcttccttgc cgcagctgtc ctcgacgttgc tctactgaagc ggaaggggac 3780
 tggctgtat tggcggaaatgc gggggggc gatcttgcgtt catctcacct tgcctctgc 3840
 gagaaggat ccatcatggc tgatgcaatgc cggccgttgc gggcacaaca gacaatcgcc 3900
 tgcccattcg accaccaagc gaaacatcgatc ctcgacgttgc cacgtactcg gatggaaagcc 3960
 ggtcttgcg atcaggatgatc tctggacgaa gagcatcagg ggctcgcc ggcacgactg 4020
 ttcgccaggc tcaaggcgcg catgcccgcg ggcgaggatc tcgtcgatc ccatggcgat 4080
 gcctgctgc cgaatatcat ggtggaaaat gggccgtttt ctggattcat cgactgtggc 4140
 cggctgggtg tggcggacccg ctatcaggac atacgcgttgc ctaccctgttga tattgtgtaa 4200
 gagcttggcg gcaatgggc tgaccgttgc ctcgtgttgc acggatcgatc cgctccccat 4260
 tcgcagcgca tcgccttgc tgcgccttgc gacgaggatc tctgagcggg actctggggat 4320
 tcgaaatgac cggcaagcg acggccaaatc tgccatcagc agatgttgc tccaccggcc 4380
 cttctatga aagggtggc ttccgatgc tttccggga cggccggctgg atgatcctcc 4440
 agcgcggggc tctcatgtcg gagttcttcg cccaccccaa cttgtttattt gcagcttata 4500
 atggttacaa ataaagcaat agcatcacaatc atttcacaatc taaagcattt ttttactgc 4560
 attcttagtttgcg tggtttgcg aaactcatca atgtatcttgc tcatgtctgt ataccgtcga 4620

cctctagcta gagcttggcg taatcatggc ctagctgtt tcctgtgtga aattgttatac 4680
 cgctcacaat tccacacaac atacgagccg gaagcataaaa gtgtaaagcc tggggtgct 4740
 aatgagttag ctaactcaca ttaattgcgt tgctgcact gcccgccttc cagtcgggaa 4800
 acctgtcggt ccagctgcat taatgaatcg gccaacgcgc ggggagaggc gttttgcggc 4860
 gagcggtatac agctcaacta aaggccgtaa tacgggttatac cacaatca ggggataacg 4920
 cagggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 4980
 tgctggcggtt ttccatagg ctccgccccctc tgacgagca tcacaatca cagcgttcc 5040
 gtcagagggtg gcgaaacccg acaggactat aaagataccg ggcgttccc cctggaaagct 5100
 ccctcggtcg ctctcctgtt ccgaccctgc cgcttaccgg atacctgtcc gccttctcc 5160
 cttcggttggc cgtggcgctt tctcaatgtc cacgtgttagt gtatctcagt tcgggttagg 5220
 tcgttcgtc caagctggc tgggtgcacg aaccggccgt tcagccgc acgttcc 5280
 tatccggtaa ctatcgctt gaggccaacc cggtaagaca cgacttatcg ccactggcag 5340
 cagccactgg taacaggatt agcagagcga ggtatgttagg cgggttaca gagttcttga 5400
 agtgggtggc taactacggc tacatagaa ggacagtggc tggtatctgc gctctgtga 5460
 acccgatttc ctccggaaaa agagttggta gcttttgcgtt gggccatcc 5520
 gtacgggtgg ttttttgcgtt tgcaaggcagc agattacgcg cagaaaaaaa gatctcaag 5580
 aagatcctt gatctttct acggggctcg acgctcgtt gAACGAAAC tcacgttaag 5640
 ggattttggc catgagatta tcaaaaaggta tcttcaccta gatccttta attaaaaat 5700
 gaagttttaa atcaatctaa agtataatag agtAAACTTG gtctgacagt taccaatgt 5760
 taatcagtga ggcacctatc tcagcgatct gtcttattcg ttcatccata gttgcctgac 5820
 tccccgtcggt gtagataact acgataacggg aggcttacc atctggcccc agtgcgtgcaa 5880
 tgataccgcg agaccacacgc tcaccggctc cagatttatac agcaataaac cagccagccg 5940
 gaaggccgca ggcagaagt ggtctgcggaa ctttatecgc cccatccatcg tcttattaaatt 6000
 gttggccgggaa agcttagatgta agtagttcg cagttaatag tttgcgcac gttgttgcca 6060
 ttgtacagg catcggtgt tcacgctcgat cggttggat ggcttcattc agctccgggt 6120
 cccaaacgatc aaggcgaggat acatgatccc ccatgttgcg caaaaaaaaagcg gtagctct 6180
 tgggtcctcc gatcggtgtc agaagtaatg tggccgcagt gttatcactc atgggtatgg 6240
 cagcaactgca taattctt actgtcatgc catccgtaaatg atgctttct gtgactggg 6300
 agtactcaac caagtccattc tgagaatatg gtatgcggcg accgagttgc tcttgcggg 6360
 cgtcaatacgc ggataatacc ggcacacata gcagaactttt aaaaatgc acatattggaa 6420
 aacgttcttc gggggcggaaaa ctctcaagga tcttaccgcgtt gttgagatcc agttcgatgt 6480
 aacccactcg tgcacccaac tgatcttcag catctttac tttcaccaggc gtttctgggt 6540
 gagcaaaaac aggaaggcaaa aatgcgcgaa aaaaggaaat aaggccgaca cggaaatgtt 6600
 gaataactcat actcttcctt tttcaatattt attgaagcat ttatcagggt tattgtctca 6660
 tgagcggata catatttgcgaa tttatggataaaaataaaca aatagggtt ccgcgcacat 6720
 ttccccggaaa agtgcacact gacgtc 6746

<210> 6
<211> 6623
<212> DNA
<213> Artificial Sequence

<220>

<220>
<223> Description of Artificial Sequence: Construct
C7LBDAS

<400> 6
gacggatcg gggatctccc gatccccat ggtcgactct cagtacaatc tgctctgtat 60
ccgcatagtt aagccagttt ctgctccctg ctttgtgttt ggaggtcgct gaggtagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcattt aagaatctgc 180
ttagggttag gcggtttgcg ctgcttcgcg atgtacgggc cagatatacg cggtgacatt 240
gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata 300
tggagttccg cggttacataa cttacggtaa atggcccgcc tggctgaccg cccaaacgacc 360
cccgcccat gacgtcaata atgacgtatg ttccatagt aacgcacata gggactttcc 420
attgacgtca atgggtggac tatttacggt aaactgccc cttggcgtt catcaagtgt 480
atcatatgcc aagtacgcgg ccttattgcgt tcaatgcgg taaatggccc gcctggcatt 540
atgcccaggta catgacccctt tgggacttcc ctacttggca gtacatctac gtattagtc 600
tcgttattac catgggtatg cggttttgcg agtacatcaa tgggctgttga tagcgggttt 660
actcacgggg atttccaatg ctccacccca ttgacgtcaa tgggagtttgc ttttggcacc 720
aaaatcaacg ggactttcca aatgtcgta acaactccgc cccatggacg caaatggccg 780
gtaggcgtgt acgggtggag gtctatataa gcagagctct ctggcttactt agagaaccca 840
ctgcttactg gcttatacgaa attaatacgat ctcactatag ggagacccaa gctggcttagc 900
gtttaaactt aagcttggta ccgagctcgat atccactatgtt ccagtgttgcgtt ggaattccctg 960
cagcccgccgg gatctatggc ccaggccggcc ctggagccctt atgcttgcggc tggcgttcc 1020
tgcgtatcgcc gcttttctaa gtcggctgtat ctgaagccat ccacacaggc 1080

cagaaggctt	tccagtgctg	aatatgcac	cgtaacttca	gtcgttagtga	ccacacctacc	1140
acccacatcc	gcacccacac	aggcgagaag	cctttgcct	gtgacatttgc	tgggaggaag	1200
tttgcagga	gtgatgaacg	caagaggcat	accaaaatcc	atthaagaca	gagggactct	1260
agaactagtt	ctgctggaga	catgagagct	gccaacctt	ggccaagccc	gctcatgatc	1320
aaacgctcta	agaagaacag	cctggcctt	tccctgacgg	ccgaccagat	ggtcagtgcc	1380
tttgttggatg	ctgagcccc	catactctat	tccgagatgt	atcctaccag	acccttcagt	1440
gaagcttcga	tgatgggctt	actgaccaac	ctggcagaca	gggagcttgt	tcacatgatc	1500
aactgggcga	agagggtgcc	aggcttgt	gattgaccc	tccatgtatc	gttccacctt	1560
ctagaatgt	cctggtaga	gatcctgtat	attgttgc	tctggcgctc	catggagcac	1620
ccagggaaagc	tactgtttgc	tcttaacttgc	ctcttggaca	ggaaccagg	aaaatgtgt	1680
gaggggcatgg	tggagatctt	cgacatgt	ctggctatc	catctcggtt	ccgcatgtat	1740
aatctgcagg	gagaggagtt	tgtgtgcctc	aaatctatta	ttttgcttaa	ttctggagtg	1800
tacacatttc	tgtccagcac	cctgaagtct	ctggaaagaga	aggaccatat	ccaccgagtc	1860
ctggacaaga	tcacagacac	tttgatccac	ctgatggcca	aggcaggcct	gaccctgcag	1920
cagcagcacc	agcggctggc	ccagctcc	ctcatcctct	cccacatcag	gcacatgagt	1980
aaccaaaggca	tggagcatct	gtacagcatg	aagtcaaga	acgtggtgcc	cctcttatgac	2040
ctgctgtctgg	agatgtgg	cggccaccc	ctacatgcgc	ccactagccg	tacgcccggc	2100
gacgcccctgg	acgacttcga	cctggacat	ctggccggcc	acgcccctgg	cgacttcgac	2160
ctggacatgc	tgcggccga	cggccctggac	gacttcgacc	tggacatgt	gcccgggtaa	2220
ctaagaatgc	ggccgctcg	gtctagagg	cccgtttaaa	cccgctgatc	agcctcgact	2280
gtgccttcta	gttgccagcc	atctgttgtt	tgcccctccc	ccgtgccttc	tttgaccctg	2340
gaaggtgcca	ctccccactgt	cctttcttaa	taaaatgagg	aaattgcata	gcattgtctg	2400
agtaggtgtc	attctattct	gggggggtgg	gtggggcagg	acagcaaggg	ggaggattgg	2460
gaagacaata	gcaggcatgc	tggggatgc	gtgggctcta	tggcttctga	ggcgaaaga	2520
accagctggg	gctctaggg	gtatccccac	gcccctgtat	gcggcgcatt	aagcgcggcg	2580
ggtgtgtgtt	ttacgcgcag	cgtgaccgct	acacttgcca	gcggccctagc	gcccgtctc	2640
ttcgcttctt	tccctcc	tctccac	ttcggccgct	ttccccgtca	agctctaaat	2700
cgggcattcc	ctttagggtt	ccgattttat	gttttacggc	acctcgaccc	caaaaaactt	2760
gatttaggtg	atggttcacg	tagtgggc	tcgcctgtat	agacggtttt	tcgcccattt	2820
acgttggagt	ccacgttctt	taatagtgg	ctcttgc	aaactggaaac	aacactcaac	2880
cctatctcg	tctattctt	tgatttataa	gggattttgg	ggatttccgc	ctattggta	2940
aaaaatgagc	tgatttaaca	aaaatttaac	gcgaattaaat	tctgtggaaat	gtgtgtcagt	3000
taggtgtgg	aaagtccca	ggctccccag	gcaggcagaa	gtatgcaaag	catgcacatc	3060
aattagtcag	caaccagg	tggaaagtcc	ccaggtccc	cagcaggcag	aagtatgcaa	3120
agcatgcac	tcaatttgc	agcaaccata	gtccccc	taactccgc	catcccgccc	3180
ctaactccgc	ccagttccgc	ccatttcc	ccccatggct	gactattttt	tttttatttt	3240
gcagaggccg	aggccgc	tgcctctg	ctattccaga	agtagtgagg	aggcttttt	3300
ggaggccctag	gcttttgc	aaagctccc	ggagcttgc	tatccatttt	cgatctgtat	3360
caagagacag	gatgaggatc	gttgc	attgaacaag	atggatttgc	cgaggttct	3420
ccggccgctt	gggtggagag	gctattcggc	tatgactgg	cacaacagac	aatcgctgc	3480
tctgtgtccg	ccgtgttccg	gctgtcagcg	caggggcgc	cggttctttt	tgtcaagacc	3540
gacctgtccg	gtgcctgaa	tgaactgc	gacgaggcag	cgccgctatc	gtggctggcc	3600
acgacggccg	ttccttgc	agctgtgtc	gacgttgtca	ctgaagcggg	aagggactgg	3660
ctgttattgg	gcaagtgcc	ggggcaggat	cttcgttcat	ctcacctgc	tcctgcccag	3720
aaagtatcca	tcatggt	tcaatgc	cggttgc	cgcttgcattc	ggctacactgc	3780
ccattcgacc	accaaggc	acatcgatc	gagcagac	gtactcgat	ggaagccgg	3840
cttgcgatc	aggatgtatc	ggacgaag	catcagg	tcgcgcac	cgaaactgttc	3900
gccaggctca	aggcgccat	gcccgcac	gaggatctc	tcgtgaccca	tggcgatgccc	3960
tgcttgcgatc	atatcatgtt	ggaaaatgc	cgctttctg	gattcatc	ctgtggccgg	4020
ctgggtgtgg	cggaccgct	tcaaggacata	gcgttggct	cccggtat	tgtgtggag	4080
cttggcggcg	aatgggtgt	ccgttcc	gtgttgc	gtatcgccgc	tcccatttc	4140
cagcgcattcg	ccttctatcg	ccttcttgc	gagtttctt	gagcgggact	ctggggatttc	4200
aaatgaccga	ccaaggc	cccaac	catcaca	tttcgtatttcc	accggccct	4260
tctatggatc	gttgggtt	ggaatcg	tccggacgc	ccgctggat	atccctccagc	4320
gccccggatct	catgctgg	ttttcgc	accccaactt	gttatttgc	gttataatgt	4380
gttacaataa	aagcaatag	atcacaat	tcacaat	agcattttt	tcactgcatt	4440
ctagttgtgg	tttgc	ctcatcaat	tatcttata	tgtctgtata	ccgtcgac	4500
ctagctagag	cttggcg	tcatggt	agcttttcc	tgtgtaaat	tgttatccgc	4560
tcacaattcc	acacaacata	cgagccgg	gcataaaat	taaaggcttgc	ggtccta	4620
gagttagtca	actcacat	attgcgttgc	gctca	cgctttcc	tcggaaacc	4680
tgtcgtgc	gtgcattaa	tgaatcg	aacgcgggg	gagaggcgtt	ttgcggcga	4740
cggatc	tcactcaa	gcccgtat	gttatcc	agaatcagg	gataacgc	4800
gaaagaacat	gtgagaaaa	ggcccgacaa	aggcaggaa	ccgtaaaaag	gcccgttgc	4860
tggcgatc	ccatagg	cgccccctt	acgacat	caaaaatc	cgctcaagtc	4920
agaggtggcg	aaacccgaca	ggactataaa	gataccaggc	gttcc	ggaagctccc	4980
tcgtgcgc	tcctgttcc	accctgc	ttaccggata	cctgtcc	tttccctt	5040
cggaaagcgt	ggcg	caatgc	gctgttaggt	tctcgttgc	gtgttaggtc	5100

ttcgctccaa gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgccgccttat 5160
 ccgtaacta tcgtcttgag tccaaccgg taagacacga cttatcgcca ctggcagcag 5220
 ccactggtaa caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt 5280
 ggtggctaa ctacggctac actagaagga cagtatttgg tatctgcgt ctgctgaagc 5340
 cagttacctt cgaaaaaaaga gttggtagct cttgatccgg caaacaaaacc accgctggta 5400
 gcggtggttt ttttggttgc aagcagcaga ttacgcgcag aaaaaaaggaa tctcaagaag 5460
 atcctttgat cttttctacg gggtctgacg ctcagtgaa cggaaaactca cgttaaggaa 5520
 ttttggtcat gagattatca aaaaggatct tcaccttagat ccttttaaat taaaatgaa 5580
 gttttaaatc aatctaaatgat atatagttt aaacttggtc tgacagttac caatgcttaa 5640
 tcagtggggc acctatctca gcgatctgtc tatttcgttc atccatagtt gcctgactcc 5700
 ccgtcggtta gataactacg atacgggagg gcttaccatc tggcccaagt gctgcaatga 5760
 taccgcgaga cccacgctca ccggctccag atttatcagc aataaaccag ccagccggaa 5820
 gggccgagcg cagaagtggt cctgcaacctt tatccgcctc catccagtct attaattgtt 5880
 gccgggaagc tagagtaagt agttcgccag ttaatagttt ggcgaacgtt gttgccattg 5940
 ctacaggcat cgtgggtgtca cgctcgctgt ttgttatggc ttcattcagc tccgggttccc 6000
 aacgatcaag gcgagttaca tgatccccca tgggtgc当地 aaaaagggtt agctcccttcg 6060
 gtcctccgtat cgttgc当地 agtaagtgg ccgcqgtgtt atcactcatg gttatggcag 6120
 cactgcataa ttcttctactt gtcatgc当地 ccgtaaagatg ctttctgtg actgggtgagt 6180
 actcaaccaa gtcattctga gaatagtgtt tgccgc当地 gagttgtct tgcccgccgt 6240
 caatacggga taataccgc当地 ccacatagca gaactttaaa agtgctcatc attggaaaac 6300
 gttcttcggg gcgaaaactc tcaaggatct taccgc当地 gagatccagt tcgatgttaac 6360
 ccactcggtc acccaactga ttctcagcat ctttacttt caccagcggt tctgggtgag 6420
 caaaaacagg aaggccaaaat gccgcaaaaa aggataaag ggcgacacgg aatgtttgaa 6480
 tactcatact cttccctttt caatattattt gaagcattt当地 tcagggttat tgtctcatga 6540
 gcgatatacat atttgaatgt atttagaaaa ataaacaaaat aggggttccg cgcacatttc 6600
 cccgaaaagt gccacctgac gtc 6623

<210> 7
 <211> 6818
 <212> DNA
 <213> Artificial Sequence

<220>

<220>
 <223> Description of Artificial Sequence: Construct
 C7LBDBL

<400> 7
 gacggatcgag gagatctccc gatccccat ggtcgactct cagtacaatc tgctctgtat 60
 ccgcatagtt aagccagtat ctgctccctg ctttgtgttt ggaggctcgct gagtagtgc当地 120
 cgagaaaaat ttaagctaca acaaggccaa gcttgaccga caatttgc当地 aagaatctgc 180
 tttagggttag gctgtttcgcc ctgc当地 atgtacggc当地 cagatatacg cgttgc当地 240
 gattatttgc当地 tagttattaa tagtaatcaa ttacggggtc attagttcat agccccatata 300
 tggagttccg cgttacataa cttacggtaa atggccgc当地 tggctgaccg cccaaacgacc 360
 cccgccccatt gacgtcaata atgacgtat ttcccatagt aacgccaata gggactttcc 420
 attgacgtca atgggtggac tatttacgtt aaactgccc当地 cttggc当地 catcaagtgt 480
 atcatatgccc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggc当地 540
 atgccc当地 catgacccta tgggactttc ctacttgc当地 gtacatctac gtatttagtca 600
 tcgctattac catggtgatg cggttttgc当地 agtacatcaa tggc当地 tggc当地 tagcgggat 660
 actcacgggg atttccaaat ctccacccca ttgacgtcaa tgggatttt当地 ttttggc当地 720
 aaaatcaacg ggacttccaa aatatgc当地 acaactccgc当地 cccatggc当地 caaatggccg 780
 gtaggcgtgt acgggtggag gtctatataa gcagagctct ctggctact agagaaccca 840
 ctgcttactg gcttatcgaa attaatacgat ctc当地 attagtgc当地 ggagacccaa gctggctagc 900
 gtttaaactt aagctggta cc当地 gagctcg当地 atccacttagt cc当地 agtgggtt当地 ggaattccctg 960
 cagccccggg gatctatggc cc当地 agggcc当地 ct当地 gagccct atgcttgc当地 tgc当地 gagtcc 1020
 tgc当地 atcgcc当地 gctttctaa gtc当地 ggctgat ct当地 gaagcgc当地 atatccgcat cc当地 acacaggc 1080
 cagaaggc当地 ttccaggctg aatatgc当地 cgtacttca gtc当地 gtagtgc当地 cc当地 accttacc 1140
 acccacatcc gc当地 accccacac aggccgagaa gctttgc当地 gtgacat当地 tgggaggaag 1200
 tttgccc当地 agtggatggc当地 caagaggccat accaaaatcc attaaagaca gagggactct 1260
 agaacttagt accggaaagg aggaggatg ttgaaacaca agcgc当地 gagag atatgtggg 1320
 gagggc当地 agtggatggg gtc当地 gctgat当地 gagatgagag ctgccaaccc ttggc当地 aagc 1380
 cc当地 gctcatgat tcaaaacgctc taagaagaac agcctggc当地 tgc当地 ctgc当地 gggccaccag 1440
 atggc当地 tagtgc当地 cttgttggc当地 tgctgagcc local cccatactct atccgagta tgatc当地 ctacc 1500
 agacccttca gtgaagctc gatgatggc ttactgacca acctggc当地 caggagctg 1560
 gttc当地 acatgatgatc当地 tcaactggc当地 gaagagggtg ccaggctt当地 tggatttgc当地 cctccatgat 1620
 caggtccacc ttctagaatg tgc当地 ctggc当地 gagatc当地 tgc当地 ttctggtct cgtctggc当地 1680

tccatggagc	acccaggaa	gctactgttt	gctcctaact	tgctcttgg	caggaaccag	1740
ggaaaatgtg	tagaggcat	ggtggagatc	ttcgacatgc	tgctggctac	atcatctcg	1800
ttccgcatga	tgaatctgca	gggagaggag	tttgtgtgcc	tcaaatactat	tattttgctt	1860
aattctggag	tgtacacatt	tctgtccagc	accctgaagt	ctctggaaaga	gaaggaccat	1920
atccaccgag	tcctggacaa	gatcacagac	actttgatcc	acctgatggc	caaggcaggc	1980
ctgaccctgc	agcagcagca	ccagcggctg	gcccgactcc	tcctcatcct	ctcccacatc	2040
aggcacatga	gtaacaaaagg	catggagcat	ctgtacagca	tgaagtgcaa	gaacgtgg	2100
cccctctatg	acctgtgtct	ggagatgtcg	gacgcccacc	gcctacatgc	gcccaactagc	2160
cgtggagggg	catccgtgga	ggagacggac	caaagccact	tggccactgc	gggctctact	2220
tcatcgcatt	ccttgaaaaa	gtattacatc	acgggggagg	cagagggttt	ccctgcccaca	2280
gtccgtatgc	cggccgacgc	cctggacgac	ttcgacctgg	acatgctgc	ggccgacgccc	2340
ctgacgact	tcgacccatg	catgtgcgg	gccgacgccc	tggacgactt	cgacctggac	2400
atgctgccgg	ggtactaag	taagcggccg	ctcgagtcta	gagggccctgt	ttaaaccgc	2460
tgatcagcct	cgactgtgc	ttctagttgc	cagccatctg	ttgtttgccc	ctccccctgt	2520
ccttccttga	ccctggaaagg	tgccactccc	actgtcctt	cctaataaaa	tgagggaaatt	2580
gcatcgcatt	gtctgagtag	gtgtcatttct	attctgggg	gtgggggtgg	gcaggacagc	2640
aagggggagg	attgggaaga	caatagcagg	catgtgggg	atgcgggtgg	ctctatggct	2700
tctgaggcgg	aaagaaccag	ctggggctct	agggggat	cccacgcgccc	ctgtagcggc	2760
gcattaagcg	cggcgggtgt	ggtgggttacg	cgacgcgtga	ccgctacact	tgccagcggc	2820
ctagcggcc	ctcccttcgc	tttcttcct	tccttctcg	ccacgttcgc	cggtttccc	2880
cgtcaagctc	taaattgggg	catcccttta	gggttccgat	ttagtgc	acggcacctc	2940
gaccccaaaaa	aacttgatta	gggtgatgtt	tcacgttagt	ggccatcgcc	ctgatagacg	3000
gtttttcgcc	ctttgacgtt	ggagtccacg	ttcttaata	gtggactctt	gttccaaact	3060
ggaacaacac	tcaaccctat	ctcggtctat	tcttttgatt	tataaggat	tttggggatt	3120
tcggcctatt	ggtaaaaaaaaa	tgagctgatt	taacaaaaat	ttaacgcgaa	ttaattctgt	3180
ggaatgtgtg	tcagttaggg	tgtggaaagt	ccccaggctc	cccaggcagg	cagaagtatg	3240
caaagcatgc	atctcaatta	gtcagcaacc	aggtgtggaa	agtccccagg	ctccccagca	3300
ggcagaagta	tgcaaaagcat	gcatctcaat	tagtcagcaa	ccatagtc	gccccctaact	3360
ccgccccatcc	cgccttaac	tccggcccaat	tccggccatt	ctccggccca	tggctacta	3420
attttttta	tttatgcaga	ggccgaggcc	gcctctgcct	ctgagctatt	ccagaagtag	3480
tgaggaggct	tttttgagg	cctaggctt	tgcaaaaaagc	tccggggagc	ttgtatatcc	3540
attttcggat	ctgatcaaga	gacaggatga	ggatcgttc	gcatgattga	acaagatgga	3600
ttgcacgcag	gttctccgc	cgcttgggt	gagaggctat	tcggctatga	ctgggcacaa	3660
cagacaatcg	gctgctctga	tgccgcccgt	ttccggctgt	cagcgcaggg	gcgcccgggt	3720
cttttgcata	agaccgacct	gtccgggtgcc	ctgaatgaac	tgcaggacgt	ggcagcgcgg	3780
ctatcggtc	tggccacgac	gggcgttct	tgcgacgtg	tgctcgacgt	tgtcaactgaa	3840
gcgggaagg	actgggtgt	attgggcgaa	gtgcggggc	agatctctt	gtcatctcac	3900
cttgctctg	ccgagaaaatg	atccatcatg	gtctgtac	tcggcggct	gcatacgc	3960
gatccggct	cctgcccatt	cgaccaccaa	gcgaaacatc	gcatcgacg	agcacgtact	4020
cggatggaaag	ccggcttctgt	cgatcaggat	gatctggacg	aagagcatca	ggggctcg	4080
ccagccgaac	tgttcgccc	gctcaaggcg	cgcatgccc	acggcgagga	tctcgctgt	4140
acccatggcg	atgcctgctt	gccgaatatc	atggggaaa	atggccgctt	ttctggattc	4200
atcgactgtg	gccggctggg	tgtggcgac	cgctatcagg	acatagcg	ggctacccgt	4260
gatattgctg	aagagcttgg	ccggcgaatgg	gctgacccgt	tcctcggt	ttacggtattc	4320
gccgtccccc	attcgacgc	catcgccctc	tatcgccctc	ttgacgatgt	cttctgagcg	4380
ggactctggg	gttgcggat	accgaccaat	cgacgcccc	cctgcattca	cgagatttc	4440
attccaccgc	cgccttctat	gaaagggttg	gcttcggat	cgtttccgg	gacgcccgg	4500
ggatgatctt	ccagcgcggg	gatctcatgc	tggagttctt	cgcccacccc	aacttggta	4560
ttgcagctt	taatggttac	aaataaagca	atagcatcac	aaatttcaca	aataaagcat	4620
tttttact	gcattctgt	tgtggtttgc	ccaaactcat	aatgtatct	tatcatgtct	4680
gtataccgtc	gacctctagc	tagagcttgc	cgtaatcatg	gtcatagct	tttcctgt	4740
gaaattgtt	tccgctcaca	attccacaca	acatacgc	cggaagcat	aagtgtaaag	4800
cctgggtgc	ctaatgatgt	agctaacttc	cattaaatgc	gttgcgtca	ctgcccggct	4860
tccactgggg	aaacctgtcg	tgccagctgc	attaatgtat	cggccaaacgc	gccccggag	4920
gcgtttcg	gcfgacggta	tcgtactact	caaaggcggt	aatacgtt	tccacagaat	4980
caggggataa	cgcaggaaag	aacatgtgag	caaaggccca	gcaaaaggcc	agaaccgt	5040
aaaaggccgc	gttgctggcg	ttttccata	ggctccgccc	ccctgacgag	catcacaaaa	5100
atcgacgctc	aagtcaaggg	ttgcgaaacc	cgacaggact	ataaaagatac	caggcg	5160
cccctggaaag	ctccctcg	cgctctcc	ttccgaccct	cccgcttacc	ggatacctgt	5220
ccgcctttct	cccttcgg	agcgtggc	tttctcaatg	ctcacgtct	agtatctca	5280
gttcgggt	ggtcggtc	tccaaactgg	gctgtgtc	cgaacccccc	gttcagcccg	5340
accgctgcgc	cttattccgt	aactatgtc	ttgagttcc	cccggtaa	cacgacttat	5400
cgcactggc	agcaggccact	ggtaaacagga	ttagcagac	gaggatgt	ggcgggtgt	5460
cagagttctt	gaagtgggt	ctaactac	gctacactag	aaggacat	tttggat	5520
gcgtctgt	gaagccagtt	accttcggaa	aaagagttgg	tagcttgc	tccggcaaac	5580
aaaccaccgc	tggtagcggt	gttttttttgc	tttgcagca	gcagattac	cgccggaaaa	5640
aaggatctca	agaagatctt	ttgatcttt	ctacgggtc	tgacgctc	tggaaacgaaa	5700

actcacgtta agggattttg gtcatagat tatcaaaaag gatcttacc tagatcctt 5760
 taaattaaaa atgaagtttt aaatcaatct aaagtataa tgagtaaact tggctcgaca 5820
 gttaccaatg cttaatcagt gaggcaccta tctcagcgat ctgtcttatt cgttcatcca 5880
 tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta ccatctggcc 5940
 ccagtgcgtc aatgataccg cgagaccac gctcacccggc tccagattt aactttatcc gcctccatcc 6000
 accagccagc cggaaaggcc gagcgcagaa gtggctctgc aactttatcc gcctccatcc 6060
 agtctattaa ttgttgcgg gaagcttagag taagtagttc gccagttaat agtttgcgc 6120
 acgttgttgc cattgtaca ggcacatgtgg tgtcacgctc gtcgttggt atggcttcat 6180
 tcagctccgg ttcccacacga tcaaggcgag ttcatgatccc cccatgttg tgcaaaaaaag 6240
 cggttagctc ctteggctt ccgatcgtt tcagaagttt gttggccgca gtgttatcac 6300
 tcatggttat ggcagactg cataatttcc ttactgtcat gccatccgtt agatgcttt 6360
 ctgtgactgg ttagtactca accaagtcat tctgagaata gtgtatgcgg cgaccgagtt 6420
 gctcttgcgc ggcgtcaata cgggataata cccgcgcaca tagcagaact ttaaaagtgc 6480
 tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat 6540
 ccagttcgat gtaacccact cgtgcaccca actgatctt acgatcttt actttccacca 6600
 gcgttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataagggcga 6660
 cacggaaatag ttgaataactc atactcttcc tttttaata ttattgaagc atttatcagg 6720
 gttattgtct catgagcgga tacatattt aatgtattt aaaaaataaaa caaatagggg 6780
 ttccgcgcac atttccccga aaagtgcac ctgacgtc 6818

<210> 8
 <211> 6695
 <212> DNA
 <213> Artificial Sequence

<220>

<220>
 <223> Description of Artificial Sequence: Construct
 C7LBDBS

<400> 8
 gacggatcgg gagatctccc gatcccattt ggtcgactct cagtacaatc tgctctgatg 60
 ccgcatagtt aagccagttt ctgctccctg ctgtgtgtt ggaggctcgct gaggtagtgcg 120
 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcgtt aagaatctgc 180
 ttagggtagt gcttttgcg ctgcttcgcg atgtacgggc cagatatacg cggtgacatt 240
 gatttttgcg tagttattaa tagtaatcaa ttacgggtt attagttcatt agcccatata 300
 tggagttccg cgttacataa ttacggtaa atggcccgcc tggctgaccg cccaaacgacc 360
 ccccccattt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420
 attgacgtca atgggttgcg tattttacggt aaactgccc cttggcgtt catcaagtgt 480
 atcatatgcc aagtacgccc cctatttgcg tcaatgacgg taaatggccc gcctggcatt 540
 atgcccagta catgacccatc tgggactttt ctacttggca gtacatctac gtatttagtca 600
 tcgcttattac catggtgatg cgggttttgcg agtacatcaa tgggcgttga tagcggtttg 660
 actcacgggg atttccaatg ctccacccca ttgacgttca tgggagttt ttttggcacc 720
 aaaatcaacg ggacttcca aatgttca acaactccgc cccatttgcg caaatgggg 780
 gttaggcgtgt acgggtggag gtctatataa cgcagactct ctggcttaactt agagaaccca 840
 ctgcttactg gcttacatcg aataatcgat ctcactatag ggagacccaa gctggcttagc 900
 gtttaaactt aagcttggta ccgagctcg atccactatgt ccagtgtgtt ggaattccctg 960
 cagcccgcccc gatctatggc ccaggcggcc ctgcggccct atgcttgcctc tgcgagttcc 1020
 tgcgatcgcc gcttttctaa gtcggctgtat ctgaagcgcc atatccgc tgcgatcgcc 1080
 cagaaggcattt tccagttgtcg aatatgcgtt cgttaacttca gtcgttagtga ccacccattacc 1140
 acccacatcc gcaccacac acggcgagaag ccttttgcct gtgacatttgc tgggagggaa 1200
 ttggccagga gtgtatgcacg caagaggcat accaaaatcc atthaagaca gagggactct 1260
 agaactatgtt accgaagagg agggagaatg ttggaaacaca agcgcacagag agatgatggg 1320
 gaggcggagg gtgaatgtgg gtctgttgcg gacatgagat ctgccaacctt ttggccaaggc 1380
 ccgctcatga tcaaacgttc taaagaagaa agcctggccct tgcgttgcg ggccgaccag 1440
 atggtcagtg cttgttgcg tgctgaggccc cccatactctt attccgatgt tgatcctacc 1500
 agacccttca gtgaagcttc gatgtatggc ttactgacca acctggcaga cagggagctg 1560
 gttcacatga tcaactggc gaagagggtt ccaggcttt tggtttgc cctccatgt 1620
 caggtccacc ttctagaatg tgcctggcta gagatcttgc tgattggctt cgtctggcc 1680
 tccatggagc acccaggaa gctactgtt gctcttactt tgctcttgcg cagggacccag 1740
 gaaaaatgtt tagaggcat ggtggagatc ttgcacatgc tgctggctac atcatctcg 1800
 ttccgcgtatc tgaatctgcg gggagggatc ttgtgtgc tcaaataatcat tattttgc 1860
 aattctggag tgtacacatt tctgtccatc accctgtatgt ctctggaaaga gaaggaccat 1920
 atccaccatgc tcctggacaa gatcacagac actttgtatcc acctgtatggc caaggcaggc 1980
 ctgaccctgc agcagcagca ccagcggctg gcccacgtcc tcctcatctt ctccccatc 2040
 aggcacatga gtaacaaagg catggagatc ctgtacatgc tgaatgttca gaacgtgggt 2100

cccctctatg acctgctgct ggagatgctg gacgcccacc gcctacatgc gcccaactagc 2160
 cgtacgccgg ccgacgcctt ggacgacttc gacctggaca tgctgcggc cgacgcctg 2220
 gacgacttcg acctggacat gctgccggcc gacgcctgg acgacttcga cctggacatg 2280
 ctgccgggt aactaagtaa gcccggcgtc gagtcttagag gccccgtta aaccgcgtga 2340
 tcagcctcga ctgtgccttc tagttgccag ccatctgtt tttgcctc ccccggtcc 2400
 tccttgaccc tggaaagggtgc cactcccact gtccttctt aataaaatga gaaaattgca 2460
 tcgcattgtc tgagtaggtg tcattctatt ctgggggggt ggggaggatt gggaaagacaa tagcaggcat gctggggatg 2520
 gaggcggaaa gaaccagctg gggctctagg gggatcccc cggtgggctc tatggcttc 2580
 ttaagcgcgg cgggtgtgt gttaacgcgc agcgtgaccg acgcgcctg tagcggcga 2640
 gccccgcctc cttecgctt ctccccttc tttctgcaca ctacacttgc cagcgcctca 2700
 caagctcaa atcggggcat ccctttaggg ttccgattt cgttgcggg ctttccccgt 2760
 cccaaaaaac ttgatttaggg ttagtgcata cgtatgggc tgcttcgcgg gtcgttacg gacactcgac 2820
 tttcgccctt tgacgttggc gtccacgttc ttaatagtg catgccttc atagacgggt 2880
 acaacactca accctatctc ggtctattt tttgatttat gcctatttgtt aaaaaatga gctgatttaa caaaatatta 2940
 atgtgtgtca gtttaggtgt gaaaagtccc caggctcccc aagggatttt ggggatttcg 3000
 agcatgcata tcaattagtc agcaaccagg tgtggaaagt acgcgaatta attctgtgga 3060
 agaagtatgc aaagcatgca tctcaattag tcagcaacca cccatccgc ccctaactcc gcccgttcc gcccattctc 3120
 ttttttattt atgcagaggc cgaggccgccc ttcgcctgtg gggaggtttt ttgcgttcc 3180
 ggaggctttt ttggaggcct aggcttttgc aaaaagctcc tagtcccgcc cctaactccog 3240
 ttccggatctg atcaagagac aggtatggaa tcgttgcga cgcggatctt gtcgactaattt 3300
 cacgcagggtt ctccggccgc ttgggtggag aggctattcg agtattcca gaagtatgtg 3360
 acaatcggtc gctctgtatgc cgccgtgttc cggctgtcag cggggagttt tatatccatt 3420
 tttgtcaaga ccgacctgtc cggtgccctt aatgaactgc tgattgaaca agatggattt 3480
 tcgtggctgg ccacgacggg cgttccttgc gcagctgtgc gctatgactg ggcacacaacag 3540
 ggaaggggact ggctgttattt gggcgaatgtt ccgcggatcc 3600
 gctctgtccg agaaaatgtt catcatgtc gatgcataatgc cgcagggttcc ttggatttcatc 3660
 ccggctacet gcccattcga ccaccaagcg aacatcgca atgaaaggccg gtcttgcgtc tcgttgcgtc 3720
 atgaaaggccg gtcttgcgtc tcaggatgtt ctggacgaaag gccgactgt cgccaggctt 3780
 gccgcgttcc agatggatcc catgttgcg ggcggctgca tacgcttgat 3840
 catggcgatg cctgcttgcg gaatatcatg gtggaaaatg tcgagcgacg acgtactcgg 3900
 gactgtggcc ggctgggtgt ggcggaccgc tatcaggaca agcatcaggg gtcgcgcgc 3960
 attgctgaag agcttggccg cgaatgggtt gaccgcttcc gcgaggatct cgtcgtgacc 4020
 gctcccgatt cgcagcgcatt cgccttctat cgccttcttg gccgcttttc ttggatttcatc 4080
 ctctgggggtt cgaatgtacc gaccaaggca cgcggatcc 4140
 ccacccgcgc ctctatgaa aggttgggt tcggaatctgc tagcgttggc tacccgtgat 4200
 tgatctcca ggcggggat ctcatgttgc agtttctgc cagcttataa ttgttacaaa taaagcaata gcatcacaaa 4260
 tttcaactgca ttcttagttgtt gtttgtcca aactcatcaa gacgcacttcc ttggatttcatc 4320
 taccgtcgac ctctatgtt agcttggcgt aatcatggtc atgatcttcc gtcgttcc 4380
 attgttatcc gtcacaaattt ccacacaaca tacgagccgg caccatccaa ttgttatttgc 4440
 ggggtgccta atgagtgagc taactcacat taattgcgtt tttcacaaat aaagcatttt 4500
 agtcggggaaa cctgtcgtgc cagctgcattt aatgaatcggt tgatcttcc 4560
 gtttgcggcc agcggatataa gctcacttca aggccgttattt atagctgtt cctgtgtgaa 4620
 gggataacgc agggaaagac atgttggca aaggccagca aagcataaaag tgtaaagcc 4680
 agggccgcgtt gctggcggtt ttccatagtc tccggcccccc gcgctcaactg cccgctttcc 4740
 gacgctcaag tcagagggtt cggaaacccca caggactata ccaacgcgcg gggagaggcg 4800
 ctgaagctc cctcgatgc ttcctgttgc cgaccctgcg acggttatcc acagaatcag 4860
 cctttctccc ttccggaaagc gtggcgctt ctcaatgttc acaggccagg aacgcgtaaaa 4920
 cggtaggt cgttcgtcc aagctgggtt gtgtgcacga tgacgacat cacaatcc 4980
 gctgcgcctt atccgttaac ttcgttgcg agtccaaacctt aagataccag gcgtttcccc 5040
 cactggcagc agccactgtt aacaggatca gcagagcgcg gcttacccga tacctgtccg 5100
 agttcttggaa gtggggccctt aactacggct acatagaaatc acgctgttagg tatctcagtt 5160
 ctctgttgcgaa gccaggatcc ttccggaaaaa gagttggtag accccccgtt cagcccgacc 5220
 ccacccgtgg tagcgggtgtt tttttgcgtt gcaagcagca ggttaagacac gacttatacgc 5280
 gatctcaaga agatcccttgc atcttttca cggggctgtca gtatgttagc ggtgtacag 5340
 cacgtttaagg gattttggcgtt atgagattt caaaaaggat gacagtattt ggtatctgcg 5400
 attaaaaatg aagttttaaa tcaatctaaa gtatatatgc ctctgttgcg gccaacaa 5460
 accaatgctt aatcagttagt gacccatctt cagcgtatctg gattacgcgc agaaaaaaa 5520
 ttgcctgtact ccccgatgc tagataacta cgatcaggca cgctcagtgg aacgaaaact 5580
 gtgtgcataa gataccgcga gacccacgtt caccggctcc gtttgcgttcc 5640
 agccagccgg aaggggccgag cgcagaatgtt gtcctgcac accatcc 5600
 ctattatgg ttccggggaa gctagatgtt gtagttgcgc tttatccgc tccatcc 5940
 ttgttgcatt tgctacaggc atcgttgcgtt caccgtcgatc agttaatagt ttgcgcac 6000
 gctccgggtt ccaacgatca aggccaggtt catgatcccc gtttggatg gcttcattca 6060
 catgttgcgca aaaaaagcgg 6120

ttagctcctt cggtcctccg atcgttgtca gaagtaagtt ggccgcagtg ttatcactca 6180
 tgtttatggc agcactgcat aattcttta ctgtcatgcc atccgtaaaga tgctttctg 6240
 tgaactggta gtactcaacc aagtcatctt gagaatagtg tatgcggcga ccgagttgct 6300
 cttggccggc gtcaatacgg gataataccg cggcacatag cagaacttta aaagtgctca 6360
 tcattggaaa acgttctcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca 6420
 gttcgatgt accccactcg gcacccaaact gatttcagc atctttact ttcaccagcg 6480
 tttctgggtg agcaaaaaca ggaaggcaaa atgcccggaaa aaaggaata agggcgacac 6540
 gggaaatgttg aataactcata ctcttcctt ttcataatatta ttgaagcatt tatcagggtt 6600
 attgtctcat gagcggatac atatttgaat gtatggaa aaataaacaat atagggggttc 6660
 cgccgcacatt tccccgaaaaa gtgccacatcg acgtc 6695

<210> 9
 <211> 6956
 <212> DNA
 <213> Artificial Sequence

<220>

<220>
 <223> Description of Artificial Sequence: Construct
 C7LBDCL

<400> 9
 gacggatcgg gagatctccc gatcccattt ggtcgactct cagtacaatc tgctctgatg 60
 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggctcgct gagtagtgcg 120
 ctagcaaaaat ttaagctaca acaaggcaag gcttgaccga caattgtcatg aagaatctgc 180
 ttaggttagtgcg ctgcttcgcg atgtacgggc cagatatacg cggtgacatt 240
 gattttagtgc tagttattaa tagtaatcaa ttacgggttc attagttcat agcccatata 300
 tggagttccg cgttacataa ttacggtaa atggccgcgc tggctgaccg cccaacgacc 360
 cccgccccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420
 attgacgtca atgggtggac tatttacggt aaactgccc cttggcagta catcaagtgt 480
 atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540
 atgcccagta catgaccta tgggactttc ctacttggca gtacatctac gtattagtca 600
 tcgcttattac catggtgatg cggttttggc agtacatcaa tggcgtgga tagcggttt 660
 actcacgggg atttccaatg ctccacccca ttgacgtcaa tggaggtttt ttttggcacc 720
 aaaatcaacg ggacttcca aatatgtcgt acaactccgc cccatgtacg ccaaattggcg 780
 gttaggctgt acgggtggag gtctatataa cgcagactct ctggcttaact agagaaccca 840
 ctgcttactg gcttatcgaa attaatacga ctcactatag ggagacccaa gctggcttag 900
 gtttaaactt aagcttggta ccgagctcg atccactagt ccagtgtgtt ggaattctg 960
 cagccccggg gatctatggc ccaggcggcc ctgcggccct atgcttgcct tgctcgagtcc 1020
 tgcgatcgcc gctttctaa gtcggctgtat ctgaagcgcc atatccgcattt ccacacaggc 1080
 cagaaggcattt tccagtgtcg aatatgcattt cgttaacttca gtcgttagtga ccacccattacc 1140
 acccacatcc gcacccacac aggccgagaag cctttgcctt gtgacatgg tgggagggaa 1200
 tttggccaggg gtatgtacgc caagaggcat accaaaatcc attaagaca gaggactct 1260
 agaacttagt gtatgttca aacataacgc tataatgtgtc cagccaccaaa ccagtgcacc 1320
 attgataaaa acaggaggaa gagctgcacg gcctggccgc tccgcacatg ctacgaatgt 1380
 ggaatgtatg aagggtggat acgaaaaagac cgaagaggag ggagaatgtt gaaacacaag 1440
 cggccagagag atgatggggaa gggcagggtt gaagtgggtt ctgctggaga catgagagct 1500
 gccaacctttt ggccaaagccc gctcatgatc aaacgcctta agaagaacacg cctggcctt 1560
 tccctgacgg cccgaccatg ggtcagtgtcc ttgttggatg ctgagcccc catactctat 1620
 tcccgagtatg atccctaccatg acccttcaatg gaagcttgcgatg tgatgggtt actgaccaac 1680
 ctggcagacca gggagcttgtt tcacatgatc aacttggcgcg agagggtgcg aggctttgt 1740
 gatggaccc tccatgatca ggttccacatc cttagatgtt cctggctaga gatcctgtatg 1800
 attggcttcg tctggcgctc ctgggacac ccaggaaatc tactgtttgc ttcttaactt 1860
 ctcttggaca ggaaccaggaa aaaatgtgtt gaggcatgg tggagatctt cgacatgtcg 1920
 ctggctacat catctcggtt ccgcacatgatc aatctgcagg gagaggatgtt tggctggctc 1980
 aaatcttattt ttttgcattaa ttcttggatgt tacacatttc tggccacac cctgaagtct 2040
 ctggaaagaga aggaccatat ccaccgcgtt ctggacaaga tcacagacac tttgatccac 2100
 ctgatggcca aggcaggcctt gaccctgcacg cagcgcacc accgcgttgc ccagctccctc 2160
 ctcatccctt cccacatcag gcacatgtt aacaaaggca tggagcatctt gtacagcatg 2220
 aagtgcagaac acgtgggtcc cctctatgac ctgtctgtgg agatgtgtt gggccacccgc 2280
 ctacatgcgc ccactagccg tggggggca tccgtggagg agacggacca aagccacttg 2340
 gccactggcg gctctacttc atgcatttc ttggaaaattt attacatcac gggggaggca 2400
 gagggtttcc ctgcccacatg ccgtacggcc gccgacgcctt tggacgactt cgacctggac 2460
 atgctggccgg cccgacccctt ggacgacttc gaccctggaca tggctggccgc cgacgcctt 2520
 gacgacttcg acctggacat gctggccggg taactaagta agcggccgtt cgagtctaga 2580
 gggcccggtt aaaccgcgtt atcagccctt actgtgcctt ctatgtgcctt ggcacatctgtt 2640

gtttccccct ccccccgtgcc ttccttgacc ctggaaagggtg ccactccac tgcctttcc 2700
 taataaaaatg aggaattgc atcgcatgt ctgagtaggt gtcattctat tctggggggt 2760
 ggggtggggc aggacagcaa gggggaggat tggaaagaca atagcaggca tgctggggat 2820
 gcgggtggct ctatggctc tgaggcggaa agaaccagct ggggctctag ggggtatccc 2880
 cacgcgcctt gtagcggcgc attaagcgcg ggggtgtgg tggtaacgcg cagcgtgacc 2940
 gctacacttgc ccagcgcctt agcgcggct ccttcgctt tcttccttc ctttcgcgc 3000
 acgttcgcgc gctttcccg tcaagctcta aatcggggca tccctttagg gttcgattt 3060
 agtgcttac ggcacactga cccaaaaaaa cttgatttagg gtatggtt acgtagtggg 3120
 ccatcgccct gatagacggt tttcgccct ttgacgttgg agtccacgtt cttaatagt 3180
 ggactcttgt tccaaactgg aacaacactc aaccctatct cggtctattt tttgattta 3240
 taagggattt tggggatttc ggcctattgg taaaaaaatg agctgattta aaaaaaattt 3300
 aacgcgaatt aattctgtgg aatgtgtgc agttaggggtg tggaaagtcc ccaggctccc 3360
 caggcaggca gaagtatgca aagcatgcat ctcaattagt cagcaaccag gtgtggaaag 3420
 tccccaggct ccccagcagg cagaagtatg caaagcatgc atctcaatta gtcagcaacc 3480
 atagtccgc ccctaactcc gccatcccg cccctaactc cggcccatatg gctgactaat ttttttatt tatcgagagg 3540
 gagetattcc agaagtagtg aggggctt tttggaggcc taggcttttgc caaaaaggtc 3600
 ccgggagctt gatatccat tttcgatcat gatcaagaga caggatgagg atcggttcg 3660
 atgattgaac aagatggatt gcacgcagg tctccggccg cttgggttga gaggctattt 3720
 ggctatgact gggcacaaca gacaatcgcc tgctctgatg cggcgttct tttgtcaag accgacctgt 3780
 ggcgcaggggc gcccgttct tttgtcaag cggcgttctt cggcgtgtca 3840
 caggacgagg cagcgcggct atcgtggctg gccacgacgg cgggtccctt gaatgaactg 3900
 ctcgacgttg tcactgaagc gggaaaggac tggctgttat gtcggatcatc 3960
 gatctccctgt catctcacct tgctcctgcc gagaaaagtat cggcggctgc atacgcttgc tccggctacc tgccatctg 4020
 atcagagcgg cacgtactcg gatggaaagcc ggtttgtcg gaggatcagg ggctcgcgcg agccaaactt ttcggcaggc 4080
 ggcgaggatc tcgtcgac ccattggcat gcctgcttc ggccgtttt ctggattcat cgactgtggc cggctgggtg 4140
 atagcgttgg ctaccgtga tattgtgaa gagcttggcg atcaggatga tctggacgaa 4200
 ctcgtcctt acggtatcgc cgctcccgat tcgcagcga cgaatatcat ggtggaaaat 4260
 gacgagttct tctgagcggg actctgggt tcgaaaatgac tggcgaccg cttatcaggac 4320
 tgcacatcgc agatttcgt tccaccgcgc ccttcattatg tttccggga cggcgttgcg atgatctcc agcgcgggg 4380
 cccaccccaa cttgttattt gcaatgttata atggttacaa ataaagcaat agcatcacaa 4440
 atttcacaaa taaagcattt ttttactgca attctagttt tccacatcg tccacacaac 4500
 atgtatctta tcatgtctgt ataccgtcga cctctagcta ctaactcaca ttaattgcgt 4560
 catagctgtt tccctgtgtga aattgttattc cgctcacaat gcaatggggc tgaccgcttc 4620
 gaagcataaa gtgtaaagcc tgggggtgcct aatgagttag gtcgctact gcccgttcc cagtcgggaa acctgtcg 4680
 gccaacgcgc ggggagggc gttttgcggc gagcggtatc ttcgttgcgc atacgtactca 4740
 tacgggttac cacagaatca ggggataacg cggaaaagaa aaaaggccag gaaccgtaaa aaggccgcgc tgctggcgtt 4800
 ctgacgagca tcacaaaaat cgacgtctaa gtcaaggggtt attcacccgt tcacccgcgc 4860
 aaagatacca ggcgttccc ctttgcggact ccctcgatcg ctttccatagg tccacacaac 4920
 cgcttaccgg atacctgtcc gccttctcc cttcggttgcg ctttccatagg tccacacaac 4980
 cacgtgttagt gtatctcagt tcgggttgcg tgcgttcgc aacagctgcgt taactacggc 5040
 aaccccccgt tcagccgcac cgctgcgcct tatccgtaa ctttccatagg tccacacaac 5100
 cggttaagaca cgacttatcg ccactggcag cagccactgg ggtatgttgcg gtcgttgcgc 5160
 ggtatgttgcg cgggtctaca gagtttcttgcg agtgggtggcc ttcgtggaaa agagttggta 5220
 ggacagtatt tggatctgc gtcgtctgcg agccaggatc ctttccatagg tccacacaac 5280
 gctcttgcgc tggcaacaa accaccgttgc tgacgggtt tttttttgtt tcaagcgcgc 5340
 agattacgcg cagaaaaaaa ggtatctcaag aagatcttgcg gatctttct acgggggtctg 5760
 acgctcgtg gaacggaaaac tcacgtttaag ggattttgtt gatctttct acgggggtctg 5820
 tcttcaccta gatccttta aattaaaaat gaagttttaa catgagatta tcaaaaaagga 5880
 agtaaaacttgc ttcgtacatgt tccatccatgc tttttttgtt tcaagcgcgc 5940
 gtcttacccatc gtcgttgcgc tccctgttgcg gatctttct acgggggtctg 5960
 aggcttacc atctggcccc agtgcgtcaca tgataccgcg gtcgttgcgc 6060
 cagatttatac agcaataaac cagccaggcg gaaggggccga ggcacccatc tcagcgatct 6120
 ctttccgcg ctccatcccg tctattatgc gtcgttgcgc 6180
 cagttaatag tttgcgcac gttgttgcga ttgttgcgc gtcgttgcgc 6240
 cgtttggat ggtttccatc agtccgggtt cccaaatgcg ttcgttgcgc 6300
 ccatgttgcg caaaaaaaaa gttagctctc tcgggttgcgc aaggcgatgtt acatgatccc 6360
 tggccgcagt gttatcactc atgggtatgg cagcaactgcg gatctttct acgttgc 6420
 catccgtaaatg atgctttctt gtgactgggtt agtactcaac taattctt acgttgc 6480
 gtatgcggcg accgagttgc tcttgcggc cgtcaatacg caagtcatc tgagaatagt 6540
 gcagaacttt aaaagtgcgtt atcattggaa aacgttcttc ggataatacc ggcacccatc 6600
 gggcgaaaaa ctctcaagga 6660

tcttaccgct gttgagatcc agttcgatgt aaccactcg tgcacccaac tgatcttcag 6720
 catctttac tttcaccagc gtttctgggt gagaaaaaac aggaaggcaa aatgccgcaa 6780
 aaaagggaaat aaggccgaca cggaaatgtt gaatactcat actcttcctt tttcaatatt 6840
 attgaagcat ttatcagggt tattgtctca tgagcggata catattgaa tgtatTTAGA 6900
 aaaataaaca aataggggtt ccgcgcacat ttccccgaaa agtgcacacct gacgtc 6956

<210> 10
<211> 6833
<212> DNA
<213> Artificial Sequence

<220>

<220>
<223> Description of Artificial Sequence: Construct
C7LBDCS

<400> 10
gacggatcg ggagatctccc gatcccstat ggtcgactct cagtacaatc tgctctgatg 60
ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120
cgagaaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatt aagaatctgc 180
ttagggttag gcgtttcg cggtttcgat atgtacgggc cagatatacg cggtgacatt 240
gattattgac tagttattaa tagtaatcaa ttacggggto attagtccat agcccatata 300
tggagttccg cggttacataa cttacggtaa atggcccccggc tggctgaccg cccaaacgacc 360
cccgccatt gacgttacataa atgacgtatg ttcccatagt aacgcataa gggactttcc 420
attgacgtca atgggtggac tatttacggt aaactgccc cttggcagta catcaagtgt 480
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540
atgcccagta catgacccta tggacttcc ctacttggca gtacatctac gtattagtca 600
tcgctattac catggtgatg cggtttggc agtacatcaa tggcgtgga tagcggtttg 660
actcacgggg atttccaagt ctccacccca ttgacgtcaa tggaggttg tttggcacc 720
aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggcg 780
gtaggcgtgt acgggtggag gtctataaa gcaagactct ctggctaact agagaaccca 840
ctgcttactg gcttatcgaa attaatacgat ctcactatag ggagacccaa gctggcttag 900
gtttaaactt aagcttggta ccgagctcgg atccactagt ccagtgttgtt ggaattccctg 960
cagcccgggg gatctatggc ccaggcgccc ctcgagccct atgcttgcct tgctcgagttc 1020
tgcgatcgcc gctttctaa gtcggctgtat ctgaagcgcc atatccgcatt cacacacaggc 1080
cagaaggcctt tccagtgtcg aatatgcatt cgtaacttca gtcgttagtgc ccaccttacc 1140
accacatcc gcaccacac aggcgagaag cctttgcct gtgacatttg tgggaggaag 1200
tttgccagga gtgatgaacg caagaggcat accaaaatcc attaagaca gaggactct 1260
agaactagta gtattcaagg acataacgac tatatgtgtc cagccaccaa ccagtgcacc 1320
attgataaaa acaggaggaa gagtcggcag gcctggcggc tccgcaatgc ttagcaagtg 1380
ggaatgatga aagggtggat acggaaagac cgaagaggag ggagaatgtt gaaacacaag 1440
cgccagagag atgatggggaa gggcagggtt gaaatgggtt ctgctggaga catgagact 1500
gccaacctt ggccaagccc gctcatgatc aaacgctcta agaagaacag cctggccttg 1560
tccctgacgg ccgaccagat ggtcagtgtc ttgttggatg ctgagccccc catactctat 1620
tccgagttatg atcctaccag acccttcaatc gaagcttgcga tgatggctt actgaccaac 1680
ctggcagaca gggagcttgt tcacatgatc aactggcga agaggggtgc aggctttgtg 1740
gatttgaccc tccatgatca ggtccacctt ctgaatgtg cctggctaga gatcctgtatg 1800
attggtctcg tctggcgctc catggagcac ccaggaaagc tactgttgc ttcttaacttg 1860
ctcttggaca ggaaccaggaa aaaaatgtta gagggtatgg tggagatctt cgacatgtg 1920
ctggctacat catctcggtt ccgcattatc aatctgcagg gagaggagtt tggatggctt 1980
aaatctatta ttttgcattaa ttctggatgt tacacatttc tgccttgcac cctgaagtt 2040
ctggaaagaga aggaccatat ccaccgagtc ctggacaaga tcacagacac ttgtatccac 2100
ctgatggcca aggccggctt gaccctgcag cagcagcacc agcggctgg ccagctccctc 2160
ctcatecctt cccacatcag gcacatgagt aacaaaggca tggagcatct gtacagcatg 2220
aagtgcaga achtgtgtcc cctctatgac ctgtgtgtgg agatgtcgaa cgcacccaccgc 2280
ctacatgcgc ccactagccg tacggccggcc gacggccctgg acgacttcga cctggacatg 2340
ctggccggccg acggccctggc cgacttcgac ctggacatgc tgccggccga cgccttggac 2400
gacttcgacc tggacatgtc gcccgggtt ctaatgttgc ggcgcgtcgatgtc tagagg 2460
cccgtttaaa cccgcgtatc agcctcgact gtgccttcta gttgcccagcc atctgttgg 2520
tggccctccc ccgtgccttc cttgaccctt gaaatgttgcga ctcccactgt cttttccctaa 2580
taaaatgagg aaattgcattt gcatgttgcg agtaggtgtc attctattct ggggggtgg 2640
gtggggcagg acagaaggaa ggaggatgg gaagacaata gcaggcatgc tggggatgac 2700
gtggggctta tggcttctga ggcggaaaga accagctggg gctctagggg gatccccac 2760
gcggccctgtt gcccgcatt aagcgccggc ggtgtggatgg ttacgcgcag cgtgaccgct 2820
acacttgcac gcccgcctac gcccgcctt ttcgccttct tcccttcctt tctgcacacg 2880
ttccgcggct ttccccgtca agctctaaat cggggcatcc ctttaggggtt ccgatttagt 2940

gctttacggc	acctcgaccc	caaaaaactt	gattagggtg	atggttcacg	tagtgccca	3000
tgcgcctgat	agacgggttt	tcgcgccttg	acgttggagt	ccacgttctt	taatagtgg	3060
cttttgtcc	aaactggaac	aacactcaac	cctatctcg	tctattctt	tgatttataa	3120
gggattttgg	ggatttcggc	ctattggta	aaaaatgagc	tgatttaaca	aaaatttaac	3180
gcgaattaat	tctgtgaaat	gtgtgtcagt	taggtgtgg	aaagtcccc	ggctccccag	3240
gcagggcagaa	gtatgcaaag	catgcatctc	aattagttag	caaccagggt	tggaaagtcc	3300
ccaggctccc	cagcaggcag	aagtatgcaa	agcatgcac	tcaatttagt	agcaaccata	3360
gtcccgcccc	taactccgc	catcccgccc	ctaactccgc	ccagttccgc	ccattctccg	3420
ccccatggct	gactaatttt	tttatttat	gcagaggccg	aggccgctc	tgccctcgag	3480
ctattccaga	agttagtgagg	aggcttttt	ggaggcctag	gctttgcaa	aaagctccc	3540
ggagcttgt	tatccatttt	cggatctgt	caagagacag	gatgaggatc	gttgcata	3600
attdaacaag	atggattgca	cgcagggtct	ccggccgctt	gggtggagag	gttattcg	3660
tatgactggg	cacaacagac	aatcggtgc	tctgtatccg	ccgtgttccg	gctgtcagcg	3720
caggggcgcc	cggttctttt	tgtcaagacc	gacctgtccg	gtgcctgaa	tgaactgcag	3780
gacgaggcag	cgcggctatc	gtggctggcc	acgacggg	ttccctgcgc	agctgtgctc	3840
gacggtgtca	ctgaaggcgg	aaggactgg	ctgttattgg	gcaagtgcc	ggggcaggat	3900
ctctgtcat	ctcacccct	tccgtcccg	aaagatcca	tcatggctga	tgcata	3960
cggctgcata	cgttgcattc	ggctacttcg	ccatcgacc	accaaggaa	acatcgatc	4020
gagcgagcac	gtactcgat	ggaagccggt	cttgcgatc	aggatgatc	ggacgaagag	4080
catcaggggc	tcgcgcagc	cgaactgttc	gccaggctca	aggcgcgc	gccgcacggc	4140
gaggatctcg	tcgtgaccca	tggcgatgcc	tgcttgcga	atatcatgtt	gaaaaatggc	4200
cgctttctg	gattcatcga	ctgtggccgg	ctgggtgtgg	cggaccgcta	tcaggacata	4260
gcgttggcta	cccggtat	tgctgaagag	cttgcggcg	aatgggctga	ccgcttcctc	4320
gtgttttacg	gtatcgccgc	tcccgattcg	cagcgcatcg	ccttctatcg	cccttcttgac	4380
gagttcttct	gagcgggact	ctggggttcg	aaatgaccga	ccaagcgac	cccaacctgc	4440
catcacgaga	tttcgattcc	accggccgc	tctatgaaag	gttgggttc	ggaatcg	4500
tccggacgc	cggctgcgt	atctcccg	cggggatc	catgttgcag	tttctcgccc	4560
accccaactt	gtttatttgc	gcttataatg	gttacaata	aagcaatagc	atcacaaatt	4620
tcacaaataa	agcattttt	tcactgcatt	ctagttgtgg	tttgcacaaa	ctcatcaatg	4680
tatcttatca	tgtctgtata	ccgtcgac	ctagctagag	cttggcgtaa	tcatggtcat	4740
agctgtttcc	tgtgtgaaat	tgttatccgc	tcacaattcc	acacaacata	cgagccggaa	4800
gcataaaatg	taaagcttgg	ggtgccta	gagttagct	actcacatta	attgcgttgc	4860
gctactgccc	cgcgttccag	tcgggaaacc	tgtcgccca	gctgcattaa	tgaatcgcc	4920
aacgcgcggg	gagaggcggt	ttgcggcgg	cggatc	tcactcaaa	gcccgtat	4980
ggttateccac	agaatcagg	gataacgc	gaaagaacat	gtgagcaaaa	ggccagcaaa	5040
agggcaggaa	cgtaaaaaa	gcccgttgc	tggcg	ccataggctc	cgcggccctg	5100
acgagcatca	caaaaatcga	cgctcaagtc	agagggtggcg	aaacccgaca	ggactataaa	5160
gataccaggc	gtttccccc	ggaagctccc	tcgtcgctc	tcctgttcc	accctgccc	5220
ttacccgata	cctgtccgc	tttctccctt	cgggaaagcg	ggcgcttct	caatgctcac	5280
gctgttaggt	tctcagg	gtgttaggtc	ttcgctccaa	gctgggtctgt	gtgcacgaa	5340
cccccggtca	ccccgaccgc	tgccgcctt	ccggtaacta	tcgtcttgag	tccaa	5400
taagacacga	cttacgc	ctggcagcag	ccactggtaa	caggattagc	agagcgaggt	5460
attagggcgg	tgctacagag	ttcttgc	ggtgcctaa	ctacggctac	actagaagga	5520
cagatattgg	tatctgcgt	ctgctgaa	cagttaac	cggaaaaa	gttggtagct	5580
tttgatccgg	caaacaacc	accgtgg	cggtgg	ttttgttgc	aagcagcaga	5640
ttacgcgcag	aaaaaaagg	tctcaaga	atcccttgc	ctttctatcg	gggtctgac	5700
ctcagtggaa	cgaaaactca	cgtttaagg	tttgc	gagattatca	aaaaggatc	5760
tcaccttagat	ccttttaat	taaaaatgaa	gtttaaatc	aatctaaatg	atatatgagt	5820
aaacttggtc	tgacagg	caatgc	tcagtggagg	acstatctca	gcgatctgc	5880
tatctgttc	atccatagtt	gcctgact	ccgtcg	gataactacg	atacgggagg	5940
gcttaccatc	tggccccc	gctgcaat	taccgc	cccacgtca	ccggctccag	6000
attatcagc	aataaaccag	ccagccgg	ggccg	cagaagtgt	cctgcaactt	6060
tatccgcctc	catccatct	attatgtt	ggcgg	tagtaagt	agttcgccag	6120
ttaatagttt	gwgcaacgtt	gttgcatt	ctacagg	cgtgtgtca	cgctcg	6180
tttgtatggc	ttcattc	tccgttccc	aacgatca	gcgagtta	tgatccccca	6240
tgttgtgca	aaaagcggtt	agtccttc	gtcctcc	cgtgtcaga	agtaagtgg	6300
ccgcagtgtt	atcactcat	gttatggc	cactgc	ttctcttact	gtcatgccc	6360
ccgtaagatg	cttttctgt	actgg	actcaacaa	gtcattctga	gaatagtgt	6420
tgccgcgacc	gagttgtct	tgcccgg	caatacgg	taataccgc	ccacatagc	6480
gaactttaaa	agtgc	atggaaa	gttctcc	gcgaaaactc	tcaaggatc	6540
taccgtgtt	gagatcc	tcgtgt	ccactcg	accaactga	tcttcagcat	6600
cttttacttt	caccagcgtt	tctgggtg	caaaaacagg	aaggcaaaat	gcccggaaaa	6660
aggaataaag	ggcgcacacgg	aaatgtt	tactcata	cttcctttt	caatattatt	6720
gaagcattta	tcagggttat	tgtctcat	gcccgtacat	atttgaatgt	atttagaaaa	6780
ataaacaaat	aggggtccg	cgcacattc	cccggaaa	gccacctgac	gtc	6833

<211> 6567
<212> DNA
<213> Artificial Sequence

<220>

<220>
<223> Description of Artificial Sequence: Construct
E2CLBDAS

<400> 11

```

gacggatcgg gagatctccc gatccctat ggtcgactct cagtaacaatc tgctctgatg 60
ccgcatatgtt aaggcagtat ctgctccctg cttgtgtgtt ggaggcgct gagtagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatt aagaatctgc 180
ttagggttag cgcttttgcg ctgcttcgcg atgtacgggc cagatatacg cggtgacatt 240
gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agccccatata 300
tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360
cccgcccat t gacgtcaata atgacgtatg ttccctatagt aacgccaata gggactttcc 420
attgacgtca atgggtggac tatttacggt aaactgccc cttggcagta catcaagtgt 480
atcatatgccc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540
atgcccagta catgacccta tgggacttgc ctacttggca gtacatctac gtattagtc 600
tcgcttattac catggtgatg cggttttgcg agtacatcaa tgggcgtgga tagcggtttg 660
actcacgggg atttccaatg ctccacccca ttgacgtcaa tgggagtttgc ttttggcacc 720
aaaatcaacg ggacttccaa aaatgtcgta aacaactccgc cccattgacg caaatggcg 780
gtaggcgtgt acgggtggag gtctatataa gcagagctct ctggcttaact agagaaccca 840
ctgcttactg gcttatcgaa attaatacga ctcaactatag ggagacccaa gctggctagc 900
gtttaaactt aagcttagat ctatggccca ggccggccctc gagccgggg agaagcccta 960
tgcttgcgg gaatgtggta agtccttctc tcagagctct cacctggcgc gccaccagcg 1020
tacccacacg ggtaaaaaac cgtataatgc cccagagtcg ggcaaatctt ttagtgactg 1080
ccgcgcaccc t gctcgccatc aacgcactca tactggcgag aagccatata aatgtccaga 1140
atgtggcaag tcttcagcc gctctgacaa gctgtgcgtt cacaacgtt ctcacacccgg 1200
taaaaaaactt agttctgtc gagacatgg agtgcacaa ctttggccaa gcccgcctat 1260
gatcaaactgc tctaaagaaga acagcctggc cttgtccctg acggccgacc agatggtcag 1320
tgccttgcgg gatgtcgatc ccccccatact ctattccggat tatgtatccta ccagaccctt 1380
cagtgaagct tcgatgtatgg gcttactgac caacctggca gacagggagc tggttcacat 1440
gatcaactgg gcgaagaggg tgccaggctt tggatgttttgc accctccatg atcaggtcc 1500
ccttctagaa tggcctggc tagagatcct gatgattggt ctcgtctggc gctccatgga 1560
gcacccaggaa aagctactgt ttgctcccaa cttgtcttgc gacaggaacc agggaaaatg 1620
ttagaggcgtt atgggtggaga tcttcgacat gctgtgcgtt acatcatctc gttccgcatt 1680
gatgaatctgg caggagggatgg agtttgcgtt cttcaatctt attattttgc ttaattctgg 1740
agtgacacata ttcttcgttca gcacccctgaa gtcgtctggaa gagaaggacc atatccaccc 1800
agtccctggac aagatcacatg acactttgtt ccaacctgtat gccaaggcag gcctgaccct 1860
gcagcagcag caccacggc tggcccagct cttccctatc ctctcccaca tcaggcacat 1920
gagtaacaaa ggcatggagc atctgtacatg catgaagtgc aagaacgtgg tggccctcta 1980
tgacctgcgt ctggagatgc tggacgccc ccccttacat gcccacta gccgtacgcc 2040
ggccgacgccc ctggacgact tcgacctggc catgtgcgcg gccgacgccc tggacgactt 2100
cgacctggac atgctccggc cccacgcctt ggacgacttc gacctggaca tggctccggg 2160
gtaaactaagt aagccggccgc tcgagtcgtt agggcccggt taaacccgtt gatcagccctc 2220
gactgtgcct tctatgtggc acccatctgt tggatgttttgc tccccctgtc cttcccttgac 2280
ccttggaggat gccactccca ctgtccttgc ctatataaaat gaggaaatttgc catgcatttgc 2340
tctgagtagg tgcattcttca ttctgggggg tgggtgggg caggacagca agggggagga 2400
ttgggaagac aatagcaggc atgctggggc tgccgtgggg tctatggctt ctgaggcgga 2460
aagaaccacgc tggggctcta gggggatcc ccaacgcgc tgcgtggcg cattaagcgc 2520
ggcgggtgtt gtgggtacgc gcagcgtac gctacactt gccagcgc gtagcgcggc 2580
tccttcgcgt ttcttcctt ctttctgcg cacttgcgc ggcttcccc gtcagctt 2640
aaatcggggc atccctttat ggttccgatt tagtgcattt cggccaccccg accccaaaaa 2700
acttgatttag ggtgtatggg cactgtatgg gccatgcgc tgcgtggcg tttttcgccc 2760
tttgacgttg gatgtcacgt tctttatag tggactttgc ttccaaactg gaacaacact 2820
caaccctatc tcggcttattt ttttggat ttttggat ttttggat cggccatttgc 2880
gttaaaaaat gagctgattt aacaaaaatt taacgcgat ttttgcgat ttttgcgat 2940
cagttagggt gtggaaagtc cccaggctcc ccaggcaggc agaagatgc aagcatgca 3000
tctcaattatc tcagcaacca ggtgtggaaa gtcggccaggc tccccctgcg gcaagatgtat 3060
gcaaagcatg catctcaattt agtcagcaac catgtcccg ccccttacatc cggccatccc 3120
gcccttaactt cccggccatc tccggccat ggtgtactaa tttttttttat 3180
ttatgcagag gccgaggccg cctctgcctc tgacttattc cagaatgtt gaggaggctt 3240
tttggaggc ttaggtttt gcaaaaatgt cccggggatc ttttgcgttca ttttgcgtt 3300
tgatcaagag acaggatgatgatcgttttgc catgattgaa caagatggat ttttgcgtt 3360

```

ttctccggcc gcttgggtgg agaggctatt cggttatgac tgggcacaac agacaatcg 3420
 ctgctctgat gcccgcgtgt tccggctgtc agcgcagggg cggccgggttc ttttgtcaa 3480
 gaccgacctg tccggtgccc tgaatgaact gcaggacgag gcagcgcggc tatcggtggct 3540
 ggccacgacg ggcgttcctt gcgcaagctgt gctcgacgtt gtcactgaag cgggaaggga 3600
 ctggctgcta ttggcgaag tgccggggca gatatctctg tcatactcacc ttgctctgc 3660
 cgagaaaagta tccatcatgg ctgatgcaat gcggcggtc catacgctt atccggctac 3720
 ctgccccattc gaccaccaag cgaaacatcg catcgagcga gcacgtactc ggtatggaagc 3780
 cggcttggc gatcgaggat atctggacga agagcatcg gggctcgcc cagccgaact 3840
 gttcgccagg ctcaaggcgc gcatgccccg cggcgaggat ctcgtctgt cccatggcga 3900
 tgcctgcttg cccaatatac tggtgaaaaa tggccgttt tctggattca tgcactgtgg 3960
 ccggctgggt gtggcggacc gctatcagga catagcggtt gctaccgtg atattgtca 4020
 agagcttggc ggcgaatggg ctgaccgtt cctcgctt tacggatatcg ccgctcccg 4080
 ttccgagcgc atcgcttct atcgcttct tgacgagttc ttctgagcgg gactctgggg 4140
 ttcaaaatga ccgaccaagc gacgcccac ctggccatcac gagattcga ttccaccggc 4200
 gccttctatg aaagggtggg cttcggaaatc gtttccggg acgcccgtg gatgatcctc 4260
 cagcgcgggg atctcatgtt gtagttctc gcccacccca acttgtttat tgcagcttat 4320
 aatggttaca aataaagca tagcatcaca aatttcacaa ataaagcatt ttttcaactg 4380
 catcttagt gtggtttgc caaaatcataatgatctt atcatgtctg tataccgtcg 4440
 acctctagct agagcttggc gtaatcatgg tcatacgctg ttccctgtgt aaattgttat 4500
 ccgctcacaa ttccacacaa catabcgacc ggaagcataa agtgtaaagc ctgggggtgcc 4560
 taatgagtga gctaactcac attaattggc ttgcgctcac tgcccgtt ccagtcggga 4620
 aacctgtcgt gccagctgca ttaatgaatc ggccaaacgcg cggggagagg cggttgcgg 4680
 cgagcggtat cagctactc aaaggcggtt atacggttat ccacagaatc aggggataac 4740
 gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg 4800
 ttgctggcgt tttccatag gtcggccccc cctgacgagc atcacaaaaa tcgacgctca 4860
 agtcagaggat ggcgaacccc gacagacta taaagatacc aggcgtttcc ccctggaaagc 4920
 tccctctgtc gctctctgt tccgaccctg cggcttaccg gataacctgtc ccctttctc 4980
 cttcgggaa gcgtggcgct ttctcaatgc tcacgctgt ggtatctcg ttccgtgttag 5040
 gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccgg ttcagccga ccgctgcgcc 5100
 ttatccggta actatcgct tgagtccaaac cggtaagac acgacttatac gccactggca 5160
 gcagccactg gtaacaggat tagcagagcg aggtatgtag gccgtgtac agagttcttg 5220
 aagtggtggc ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgtcg 5280
 aaggcagtttta ctttcggaaa aagagtgtt agctcttgc gtcggccaa aaccaccgc 5340
 ggtagcgggt gttttttgt ttgcaagcag cagattacgc gcagaaaaaaa agatctcaa 5400
 gaagatcctt tgatcttttacgggtgtct gacgctcgtt ggaacggaaaaa ctcacgttaa 5460
 gggattttgg tcatgagatt atcaaaaaagg atcttcaccc agatectttt aaattaaaaa 5520
 tgaagttta aatcaatcta aagtatataat ggttaactt ggtctgacag ttaccaatgc 5580
 ttaatcagtg aggcacctat ctcagcgatc tgtctatttc gttcatccat agttgcctga 5640
 ctccccgtcg ttagataaac tacgataacgg gagggtttac catctggccc cagtgcgtca 5700
 atgataccgc gagacccacg ctcaccggct ccagatttat cagcaataaa ccagccagcc 5760
 ggaaggggccg agcgcagaag tggctctgca actttatccg cttccatcca gtctattaat 5820
 tggccgggg aagctagatg aagttagttcg ccagttataa gtttgcgcaaa cgttggcgc 5880
 attgctacag gcatcggtt gtcacgctcg tggcttggta tggcttcatt cagctccgg 5940
 tcccaacgat caaggcgagt tacatgatcc cccatgtgt gcaaaaaaagc gtttagctcc 6000
 ttccggtcctc cgatcggtt cagaagtaat ttggcccgag tggatctact catggttatg 6060
 gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgctttc tggactgg 6120
 gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg 6180
 gcgtcaatac gggataatac cgcgcacat agcagaactt taaaagtgtct catcattgg 6240
 aaacgttctt cggggcgaaa actctcaagg atcttaccgc tggtagatc cagtgcgtat 6300
 taacccactc gtgcacccaa ctgatctca gcatttta ctttcaccag cgttctggg 6360
 ttagcaaaaaa caggaaggca aaatgcgcgaa aaaaaggaa taagggcgcac acggaaatgt 6420
 tgaatactca tactcttctt tttcaatataat tattgaagca ttatcaggg ttattgtctc 6480
 atgagcggat acatatttga atgtatattag aaaaataaaac aaataggggt tccgcgcaca 6540
 ttcccccggaa aagtgcacc tgacgtc 6567

<210> 12
 <211> 6639
 <212> DNA
 <213> Artificial Sequence

<220>

<220>
 <223> Description of Artificial Sequence: Construct
 E2CLBDBS

<400> 12

gacggatcg gagatctccc gatccccat ggtcgactct cagtacaatc tgctctgatg 60
 ccgcatagtt aaggccatgt ctgctccctg ctttgtgttt ggaggtcgct gagtagtgcg 120
 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatt aagaatctgc 180
 ttagggtagt gcgtttcg cgttcataa atgtacgggc cagatatacg cggtgacatt 240
 gattattgac tagtattaa tagtaatcaa ttacggggc attagttcat agccatata 300
 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360
 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420
 attgacgtca atgggtggac tatttacgt aaactgccc cttggcagta catcaagtgt 480
 atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540
 atgcccagta catgacccctt tgggactttc ctacttggca gtacatctac gtattagtca 600
 tcgctattac catggtgatg cgggtttggc agtacatcaa tgggctgga tagcggttt 660
 actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttgg ttttggcacc 720
 aaaatcaacg ggacttcca aaatgtcgta acaactccgc cccattgacg caaatggcg 780
 gtagggcgtgt acgggtggag gtctatataa gcagagctct ctggctaact agagaacca 840
 ctgcttaactg gcttatcgaa attaatcgaa ctcactatag ggagacccaa gctggcttag 900
 gtttaaactt aagcttagat ctatggccca ggcggccctt gagccgggg agaagcccta 960
 tgcttgcgg gaatgtggta agtccttctc tcagagctct cacctggc gccaccagcg 1020
 taccacacg ggtaaaaaac cgtataatgc cccagagtgc ggc当地atctt ttagtgcactg 1080
 cccgacactt gctcccccatt aacgcactca tactggcgag aagccataca aatgtccaga 1140
 atgtggcaag tctttcagcc gctctgacaa gctggcgctt caccacgtta ctcacaccgg 1200
 taaaaaaaaact agtgaccgaa gaggaggag aattttggaa cacaagcgcc agagagatga 1260
 tggggagggc aggggtgaag tggggctgtc tggagacatg agagctgcca acctttggcc 1320
 aagccgctc atgatcaaac gctctaagaa gaacagcctg gccttgcgg tgcggccga 1380
 ccagatggtc agtgcctgt tggatgtcg gccccccata ctcttcccg agtatgatcc 1440
 taccagacc ttcatgtaaat cttcgatgtat gggctactt accaaccctgg cagacaggga 1500
 gctgggtcac atgatcaact gggcaagag ggtggccaggc tttgtggatt tgaccctcca 1560
 tgcgtcggc caccttctatg aatgtgcctg gctagagatc ctgtatgtt gtcgtctg 1620
 ggc当地ccatg gagcaccctt ggaagctact gtttgcctt aacttgcctt tggacaggaa 1680
 ccaggaaaat tttgttaggg gcatggtgg gatttcgac atgctgctgg ctacatcatc 1740
 tcgggtccgc atgatgaatc tgcaggggaga ggagtttgg tgcctcaat ctattattt 1800
 gcttaattct ggagtgtaca catttctgtc cagcaccctt aagtctctgg aagagaagga 1860
 ccatatccac cgagtcctgg acaagatcac agacactttt atccacctga tggccaaaggc 1920
 aggctgacc ctgcagcgc agcaccagcg gtcggccctt ctcccttca tcctctccca 1980
 catcaggcac atgatgaaata aaggcatgg gcatctgtac agctgatgtt gcaagaacgt 2040
 ggtggccctc tatgacctgc tgctggagat gtcggacgca caccgcctac atgcgccccac 2100
 tagccgtacg cggccgacg ccctggacga cttcgacctg gacatgctgc cggccgacgc 2160
 cctggacgac ttgcacctgg acatgctggc ggccgacgca ctggacgact tcgaccttgg 2220
 catgctccgg gggtaactaa gtaagcgcc gtcgagttt agagggcccg ttaaaccctgg 2280
 ctgatcagcc tcgactgtgc cttctagttt ccagccatct gttgtggcc ctcctccctgt 2340
 gccttccttgc accctggaaat gtggccactcc cactgtctt tccataaaa atgaggaat 2400
 tgcgtcgtat tttctgttgc ggtgtcattt tattctgggg ggtgggtgg ggcaggacag 2460
 caagggggg gattgggaag acaatagcag gcatgtctttt gatcggtgg gctctatggc 2520
 ttctggggc gaaagaacca gtcggggctc taggggttat ccccacgcgc cctgttagcgg 2580
 cgcattaagc gcgccgggtg tgggtggttac ggc当地ggcgtg accgctacac ttgcccagcgc 2640
 cctagcgccttgc ctcccttcc ctttcttccc ttcccttctc gccacgttcg cccgctttcc 2700
 ccgtcaagct ctaaatcgaa gcatccctt agggtccga tttagtgttt tacggcacct 2760
 cgaccccaaa aaacttgatt agggtgtatgg ttcaatgtatc gggccatcgcc ctcgtatagac 2820
 ggttttcgc ctcccttgc tggagtcac gttctttat agtggactct tttccaaac 2880
 tggacaacaat ctcaacccctt ttcctgtcta ttctttgtat ttataaggaa ttttggggat 2940
 ttccggcttgc tggttaaaaat atgagctgtat ttaacaaaaat ttaacgcga attaattctg 3000
 tggaaatgtgt gtcaggtagg gttgtggaaat tccctggat ccccaggcg gcaagaatgtat 3060
 gcaaaagcatg catctcaattt agtcagcaac caggtgtggaa aagtcccttgc gtcggccacgc 3120
 aggccagaatg atgcaaaagca tgcataatc ttagtcagca accatagtcc cggccctaaac 3180
 tccggccatc cggcccttgc ctccggccatc ttccggccatc tctccggcccc atggctgact 3240
 aattttttt atttatgcac aggccgaggc cgccatgtcc tctgagctat tccagaagata 3300
 gtgaggaggc ttttttggag gcctaggctt ttgcaaaaat cttccggggat ctcggggatc 3360
 cattttcgatc tctgtatcaatc agacaggatg aggatgtttt cgc当地tgcattt aacaagatgg 3420
 attgcacgcgca ggttctccgg cgc当地tgggtt ggaagggtca ttcggctatg actggggcaca 3480
 acagacaatc ggctgtctgc atgcccgcgt gttccggctt ctagcgcagg ggc当地ggcgtt 3540
 tcttttgc aagaccgacc tgc当地ggcgttcc cttcaatgtaa ctgcaggacg aggccggccgc 3600
 gctatcgtgg ctggccacga cggccgttcc ttgc当地ggcgtt gtc当地ggcgtt ttgtcactga 3660
 agcgggaagg gactggctgc tattttggcga agtgc当地ggggc caggatctcc tgc当地tgcattt 3720
 ctttgccttgc gccgagaaat tatccatcat ggc当地gtatgc atgc当地ggcgc tgc当地tgcattt 3780
 tgc当地ggcttgc acctggccatc tcgaccacca agc当地aaatcatc cgc当地tgcattt gacggccagg 3840
 tcggatggaa gccggcttgc tcgatcagga tgc当地tggac gacggccatc gggggctcgc 3900
 gccagccgaa ctgttgc当地 ggc当地tggac ggc当地atgc gacggccagg attcgtctgt 3960
 gacccatggc gatgc当地tgc当地 tgc当地tggatcatc gatgggtggaa aatggccgct tttctggatt 4020

catcgactgt ggccggctgg gtgtggcgaa ccgcgtatcg gacatagcgt tggctacccg 4080
tgcatttgcg gaagagcttg gcggcgaatg ggctgaccgc ttccctcgatc tttacggat 4140
cgccgctccc gattcgcagc gcatgcctt ctatgcctt cttgacgagt tcttctgagc 4200
gggactctgg ggttcgaaat gaccgaccaa gcgaacccca acctgccatc acgagatttc 4260
gattccaccg ccgccttcta taaaagggttgg ggcctcgaa tcgtttccg ggacgcccgc 4320
tggatgatcc tccagcgcgg ggatctcatg ctggagttct tcgcccaccc caacttgc 4380
attgcagctt ataatgtta caaaataaagc aatagcatca caaatttcac aaataaaagca 4440
ttttttcac tgcattcttag ttgtggtttg tccaaactca tcaatgtatc ttatcatgtc 4500
tgataaccgt cgaccccttag cttagatcg gctaattatc ggtcatagct gtttctgtc 4560
tgaaaattgtt atccgctcac aattccacac aacatacggg ccggaagcat aaagtgtaaa 4620
gcctgggggtg cctaattgagt gagctaactc acattaatttgcgttgc gtcggcaacg cgccggggaga 4680
ttccagtcgg gaaacctgtc gtgcagctg catatgaa taatacgggtt atccacagaa 4800
ggcggttgc ggcgagcggg atcagctcac tcaaaggcg gtcacgctg taggtatctc 4860
tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc cccctgacga gcatcacaaa 4920
aaaaaggccg cggtgcgtgc gtttttccat aggctccgcg tataaagata ccaggcggtt 4980
aatcgacgtt caagtcagag gtggcgaaac ccgacaggac tgccgttac cgatatactg 5040
ccccctggaa gtccttcgt ggcctctctt gttccgaccc gtcacgctg taggtatctc 5100
tccgccttc tcccttcggg aacgtgggg ctttctcaat acgacccccc cggtcagccc 5160
agttcgggtt aggtcgttcg ctccaagctg ggctgtgtc accccggtaag acacgactta 5220
gaccgctcgc ctttatccgg taactatctg cttgagtcgaa cgaggtatgt aggccgtgtc 5280
tcgcccactgg cagcagccac tggtaacagg attagcagag gaaggacagt atttggtatac 5340
acagagttct tgaagtgggt gcttaactac ggctacacta tggctcttgcg atccggcaaa 5400
tgcgtctgc tgaagccagt taccttcgaa aaaagagtttgcg acgagattac ggcagaaaa 5460
caaaccacccg ctggtagcgg tggttttttt gtttgcagc ctgacgtca gtggaaacgaa 5520
aaaggatctc aagaagatcc ttgtatctt tctatgggggtt ggtatcttac ctatgtt 5580
aactcacgtt aagggtttt ggtcatgaga ttatcaaaaa atgataaac ttgttctgac 5640
ttaaattaaa aatgaattttt taaatcaatc taaatgtat tctgtcttatt tcgttcatcc 5700
agttaccaat gcttaatcg tgagggcacct atctcagcga gggagggctt accatctggc 5760
atagttgcct gactcccggt cgtgtagata actacgatac ctccagattt atcagcaata 5820
cccagtgcgt caatgatacc gcgagaccca cgctcaccgg caactttatc cgccctccatc 5880
aaccagccag ccgaaaggccg cgagcgcaga agtggctctg cgcgcgttaa tagttgcgc 5940
cagtctatta atgttgcgg ggaagctaga gtaagtatgtt cgtcggttgg tatggcttca 6000
aacgttgggtt ccattgtac aggcatcggt gtgtcaacgtt ccccatgtt gtgcaaaaaaaa 6060
ttcagctccg gttcccaacg atcaaggcga gttcatgtat agtggccgc agtgttatca 6120
gcggttagt cttcgggtcc tccgatcggtt gtcagaagta tgccatccgt aagatgtttt 6180
ctcatggta tggcagcaact gctataatttctt cttactgtca agtgtatgcg gcgaccgagt 6240
tctgtgactg gtgagtaactc aaccaagtcg ttctgagaat atagcagaac ttaaaatgt 6300
tgctcttgcg cggcgtcaat acgggataat accggccac ggtatctacc gctgttgaga 6360
ctcatcattt gaaaacgttc ttccccggcga aaactctcaa cagcatcttt tactttcacc 6420
tccagttcga tggtaacccac tcgtgcaccc aactgatctt aaaaaagg aataaggcg 6480
agcgttctg ggtgagcaaa aacaggaagg caaaatgccg acatgttacatcatttcaat 6540
acacggaaat gttgaataact catactcttc tttttcaat attattgaag cattttatcag 6600
ggttattgtc tcatgagcgg atacatattt gaaatgttattt agaaaaataa acaaataagg 6660
gttccgcgca catttccccgg aaaaatgcga cctgacgtc 6639

<210> 13
<211> 6801
<212> DNA
<213> Artificial Sequence

<220>

<220>
<223> Description of Artificial Sequence: Construct
LBDASNLSVP16

<400> 13
gacggatcgg gagatctccc gatccctat ggtcgactct cagtcataatc tgctctgtatc 60
ccgcatacgat taaaggatctt ctgtccctgt cttgtgttgg ggaggtcgct gaggtagtgcg 120
cgagcaaat ttaagctaca acaaggcgaag gcttgcacca caattgcgtt aagaatctgc 180
ttagggttag gcggttgcg ctgcgttcgg atgtacgggc cagatatacg cgttgcacatt 240
gattattgtac tagttattaa tagtaatcaa ttacgggttc attagttcat agcccatata 300
tggagttccg cgttacataa cttacggtaa atggccgcgc tggctgaccg cccaacgacc 360
cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgcataa gggactttcc 420
atggacgtca atgggtggac tatttacgtt aaactgcaca cttggcagta catcaagtgt 480
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540
atgcccagta catgacccatc tgggactttc ctacttggca gtacatctac gtattagtca 600

tcgctattac catggtgatg cggtttggc agtacatcaa tgggcgtgga tagcggtttg 660
 actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttgc ttttggcacc 720
 aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg 780
 gtaggcgtgt acgggtggag gtctataaa gcaagagctc ctggctaact agagaacca 840
 ctgcttactg gcttatcgaa attaatcgaa ctcactatag ggagacccaa gctggctagc 900
 gtttaaactt aagcttggta ccagagctcg attcaactatg ccagtggtt ggaattccctg 960
 cagccccggg gatctatggc ccaggcgccc ctcgagccct atgctgccc tgctcgatgtc 1020
 tgcgatcgcc gctttctaa gtcggctgt ctgaagcgcc atatccgcat ccacacaggc 1080
 cagaaggcctt tccagtgtcg aatatgcattg cgtaaacttca gtcgtagtga ccaccttacc 1140
 acccacatcc gcaccacac aggcgagaag cctttgcct gtgacatttgc tgggaggaag 1200
 tttgccagga gtgatgaacg caagaggcat accaaaatcc atttaagaca gagggactct 1260
 agaacttagt ctgctggaga catgagagct gccaacctt gccaagccc gctcatgatc 1320
 aaacgctcta agaagaacag cctggccttgc tccctgacgg ccgaccatg ggtcagtgcc 1380
 ttgttggatg ctgagcccc catactatc tccgagatgtc atccattaccatcccttca 1440
 gaagcttcga tgatggcctt actgaccaac ctggcagaca gggagctgtt tcacatgatc 1500
 aactggcga agagggtgccc aggcttgc gattgaccc tccatgatca gttccaccc 1560
 ctagaatgtg cctggctaga gatcctgatg atttgtctcg tctggcgctc catggagcac 1620
 ccagggaagc tactgtttgc tcctaacttgc tctttggaca ggaaccaggaaaaatgtgta 1680
 gagggcattgg tggagatctt cgacatgtcg ctggctacat catctcggtt ccgcatgatg 1740
 aatctgcagg gagaggagtt tttgtgcctc aaatcttata ttttgcattaa ttctggagtg 1800
 tacacatttc tgtccagcac cctgaagtct ctgaaagaga aggaccatcatcccgatgc 1860
 ctggacaaga tcacagacac tttgtatccatc ctgatggccca aggccggctt gaccctgcag 1920
 cagcagcaccc agccgttgcg ccagctctc ctatctctc cccacatcag gcacatgatg 1980
 aacaaaggca tggagcatct gtacagcatg aagtgcagaacgctgggtccctatgac 2040
 ctgctgtcg agatgtcgatc cggccacccgc ctacatgcgc ccactagccg tacgcccggaa 2100
 aagaaacgcg aagttggcg cggccggcgct ccccccggccatgatcccttgcgactgt 2160
 ctccacttag acggcgagga cgtggcgatg ggcgcattgc acgcgtatgatccatcg 2220
 ctggacatgt tggggacgg ggattccccg ggtccggat ttaccccccgcgactccg 2280
 ccctacggcg ctctggatat ggccgacttc gagtttgcg agatgtttac cgatgcccctt 2340
 ggaattgacg agtacggttt aattaactac cctgtacgacg ttccggactatcgcttgc 2400
 gaattcgcgg cgcgtcgatc cttagggcc cgtttaaacccctgtatcgactgt 2460
 gccttcttagt tgccagccat ctgttgcggcccttgcgttccctgtacccttgc 2520
 aggtggcact cccactgtcc tttcttaataa aataggagaa attgcattgcg attgtctgag 2580
 taggtgtcat tctattctgg ggggtgggtt gggcaggac agcaaggggg aggattggaa 2640
 agacaatagc aggcatgtcg gggatgcggatc gggctctatg gcttctgagg cggaaagaac 2700
 cagctggggc tctaggggatc atccccacgc gcccgtatgc ggcgcattaa ggcggccggg 2760
 tgggtgggtt acgcgcagcg tgaccgtac acttgcgcg gcccctagcgc ccgctccctt 2820
 cgctttcttc ctttcctttc tcgcccacgtt cgcggcttttccctgtatcg 2880
 gggcatccct ttagggttcc gatttagtgc ttacggcactcgcggccatc aaaaacttgc 2940
 ttaggggtat ggttacgttgc gttggccatc gcccgtatgc acggttttccgccttttgc 3000
 gttggagtc acgttcttta atagtggact ctgttccaa actggaaacaa cactcaaccc 3060
 tatctcggtc tatttttttgcattataagg gattttgggg atttcggcctt attggtaaaa 3120
 aaatgagctg attaacaaa aattnaactgc gaattaatttgc ttttttttttgc 3180
 ggggtgtggaa agtccccagg ctccccaggc aggcagaatgttgcattctca 3240
 ttagtcagca accaggtgtg gaaagtccccca aggtccccca gcaggcagaa gtatgc 3300
 catgcatttc aatttagtcag caaccatagt cccccccta actccgcacccatcc 3360
 aactccgccc agtccggccc attctccgccc ccatggctga ctaatttttttgc 3420
 agaggcccgag gcccctctg cctctgatc attccagaag tagtgaggag gcttttttgc 3480
 aggcttaggc ttttgcaaaaa agtccccccgggg agcttgcatttttcgcgatctgatca 3540
 agagacagga tgaggatgttgc ttcgcattatgc tggatcaatgcgatgttgc 3600
 gggccgttgg gttggagggc tattcggctatgc tgactggca caacagacaa tggctgtctc 3660
 tgatggccgc gtgttccggc ttttcggctatgc gggccggcccg gtttttttgc tcaagacccg 3720
 cctgtccggc gcccgtatgc aactgcggatc cgaggcagcg cggctatcgatgc 3780
 gacggccgtt ctttcggccatc ctgttgcgtatgc cgttgcacttgc gaaaggccaa gggactggct 3840
 gctattgggc gaagtggccgg ggcaggatct cctgtatgc cacctgtctc tggccgagaa 3900
 agtattccatc atgggtgtatc caatgcggcg gtcgcatacg cttgtatccgg ctacctgc 3960
 attcgaccac caagcgaaac atcgcatcgatc gcgaggcacttgc acgtggatgg aagccggct 4020
 tggcgatcgatc gatgtatcgatc gcaagagatc tcaggggctc gcccggccatc aactgttgc 4080
 caggctcaag ggcgcgtatgc cggacggcgatc ggcgcgtatgc gtcgatgcctg 4140
 cttggccatc atcatgggttgc aaaaatggccg cttttctggatc ttcatcgact tggccggct 4200
 ggggtgtggcg gaccgtatc aggacatgc gttggctacc cgtgatatttgc tggatgc 4260
 tggcgccgaa tgggtgtacc gtttcctcgatc gttttacggatc atcgccgtatgc cccattcgatc 4320
 ggcgcattcgcc ttctatcgatc ttcttgcgtatgc gttttctgc gcccggactatgc ggggttgc 4380
 atgaccgacc aagcgacgccc caacctgcata tcacgagatc tcgatccac cgcggcccttc 4440
 tatgaaaggt tgggttccgg aatcgatccatc cggacggccatc gtcggatgtatgc cttccagcg 4500
 gggatctca tggctggatgttgc cttcgccatc cccaaacttgc ttattgcagc ttataatgtt 4560
 tacaataaaa gcaatagcat cacaatatttgc acaaaatggccatcatttttgc actgcatttc 4620

agttgtggtt tgccttcaact catcaatgtc tcttatcatg tctgtataacc gtcgacacct 4680
 agctagagct tggcgtaatc atggtcatacg ctgtttccctg tgtgaatttgc ttatccgctc 4740
 acaattccac acaacatacg agccggaagc ataaagtgtc aaggctgggg tgccataatgtc 4800
 gtgagctaac tcacattaat tgcgttgcgc tcactgccc ctttccagtc gggaaacctg 4860
 tcgtgccagc tgcattaatg aatcgccaa cgccggggg gaggcggtt gcggcgagcg 4920
 gtatcagtc actcaaaaggc gtaatacagg ttatccacag aatcaggaa taacgcagga 4980
 aagaacatgt gagcaaaaagg ccagcaaaag gccaggaaacc gtaaaaaggc cgccgttgctg 5040
 gcgttttcc ataggctcg cccccctgc gaggatcaca aaaatcgacg ctcaagttag 5100
 aggtggcgaa acccgacagg actataaaa taccaggct ttccccctgg aagctccctc 5160
 gtgcgtctc ctgttccgac cctgcccgtt accggatacc tgcgttccctt tctcccttcg 5220
 ggaagcgtgg cgctttctca atgctcacgc tgcgttccgtt tcagttcggt gttaggtcg 5280
 cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc 5340
 ggtaactatc gtcttgcgtc caacccggta agacacgact tatcgccact ggcagcagcc 5400
 actggtaaca ggatttagcag agcgaggatgt gtaggcgggt ctacagagtt ctgttgtgg 5460
 tggcctaact acggctacac tagaaggaca gtatggta tctgcgtctc gctgaagcca 5520
 gttaccttcg gaaaaagagt tggtagctt tgatccggca aacaaaccac cgctggtagc 5580
 ggtgggtttt ttgttgcggc gcagcggat acgcgcgaa aaaaaggatc tcaagaagat 5640
 ccttgcgtatct ttgttgcggg gtctgcgtc cagttggaaacg aaaatcagc ttaaggggatt 5700
 ttgttgcgtatc gattatcaa aaggatctt accttgcgtt ttttaattttaaaaatgaagt 5760
 tttaaatcaa tctaaagtat atatgatcaa acttgcgttcat acagttacca atgcttaatc 5820
 agtgaggcac ctatctcagc gatctgttca tttgcgttcat ccatagttgc ctgactcccc 5880
 gtcgtgtaga taactacgt acgggaggcc ttaccatctg gcccagtgc tgcaatgata 5940
 cccgcgagacc cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg 6000
 gcccggcgca gaagtggtcc tgcaacttta tccgcctcca tccagtttat taattgttgc 6060
 cgggaagctt gagaatgtt ctgcgttgcgc aatagggttc gcaacgttgc tgccattgtc 6120
 acaggcgtcg tgggtgcacg ctgcgttgcgtt ggtatggcc cattcagtc cggttcccaa 6180
 cgatcaaggc gaggatcatg atccccatg ttgttgcgggg aagcgggttag ctcccttcgg 6240
 cctccgatcg ttgttgcggg taagttggcc gcagtttat cactcatgg tatggcagca 6300
 ctgcataatt ctcttactgt catgcatttc gtaagatgtt tttctgtgac tggtgagttac 6360
 tcaaccaagt cattctgaga atagtgatgtt cggcgaccga gttgcttgc cccggcggtca 6420
 atacgggata ataccgcgc acatagcaga actttaaaatg tgctcatcat tgaaaacgt 6480
 tcttcggggc gaaaactctc aaggatctt ccgcgttgc gatccagttc gatgttaaccc 6540
 actctgtcac ccaactgttca ttcagcatct tttactttca ccagcgtttc tggttgagca 6600
 aaaacaggaa ggcaaaatgc cgcaaaaaaaag ggaataagggg cgacacggaa atgttgaata 6660
 ctctactct tccttttca atttatttgc agcatttttgc aggttatttgc ttcatgagc 6720
 ggatacatat ttgaatgtt attagaaaaat aaacaaatag gggttccgcg cacatttcccc 6780
 cgaaaagtgc cacctgacgt c 6801

<210> 14
 <211> 6695
 <212> DNA
 <213> Artificial Sequence

<220>

<220>
 <223> Description of Artificial Sequence: Construct
 LBDBSG400V

<400> 14
 gacggatcg gaggatctccc gatcccttat ggtcgactct cagtcataatc tgctctgtatg 60
 ccgcataatc aagccgttat ctgcgttccctg cttgtgtgtt ggagggtcgct gaggatgtcg 120
 cgagcaaaat ttaagactaca acaaggcaag gcttgcaccg caattgcatg aagaatctgc 180
 ttaggggttag gcgttttgcg ctgcgttgcg atgtacgggc cagatatacg cggtgacatt 240
 gattatttgc tagttattaa tagtaatcaa ttacgggttc attagttcat agcccatata 300
 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360
 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgcacaata gggactttcc 420
 attgacgtca atgggtggac tatttgcgtt aaactgcaca cttggcagta catcaagtgt 480
 atccatatgcc aagtacgccc cttattgacg tcaatgacgg taaatggccc gcctggcatt 540
 atgcccgtt catgacccca tgggactttt ctacttgcgca gtacatctac gtattagtca 600
 tcgttattac catgggtatg cgggttttgc agtacatcaa tgggcgttgc tagcggtttg 660
 actcacgggg atttccaaatg ctccacccca ttgcgttgc tggagtttgc ttttggcacc 720
 aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggcg 780
 gtagggcgtgt acgggtggag gtcttatataa gcagagctct ctggcttaact agagaaccca 840
 ctgttacttgc gcttatacgaa attaatacgat ctcactatag ggagacccaa gctggcttagc 900
 gtttaaactt aagcttgcgtt ccggacgttgcg atccactatgt ccagtgttgc ggaattccctg 960
 cagcccgcccc gatctatggc ccaggccggcc ctgcggccct atgcttgcggc tgcgttgc 1020

tgcgatcgcc gctttctaa gtcggctgat ctgaagcgcc atatccgcac ccacacaggc 1080
 cagaaggcctt tccagtgtcg aatatgcatt cgtaacttca gtcgttagtga ccaccttacc 1140
 acccacatcc gcaccacac acggcagaag cctttgcct gtgacatttgg tgggaggaag 1200
 tttgccagga gtatgaaacg caagaggcat accaaaatcc attaaagaca gagggactt 1260
 agaactagtg accgaagagg acggagaatg ttgaaacaca agcgcacagag agatgatggg 1320
 gagggcagggt gtgaatgtgg gtctgctgg gacatgagag ctgccaacct ttggccaagc 1380
 cccgtcatga tcaaactgctc taagaagaac agctggct tgccttgc ggcgcaccag 1440
 atggctagtg ctttggatgg tgctgagccc cccatactct attccgagta tgatcctacc 1500
 agacccttca gtgaagcttc gatgatgggc ttactgacca acctggcaga cagggagctg 1560
 gttcacatga tcaactgggc gaagaggggt ccaggctttg tggatttgac cctccatgat 1620
 caggtccacc ttctagaatg tgcctggta gagatcctga tgattggct cgtctggcgc 1680
 tccatggagc acccagtgaa gctactgttt gctctaact tgctcttggc caggaaccag 1740
 ggaaaatgtg tagagggcat ggtggagatc ttcgacatgc tgctggctac atcatctcg 1800
 ttccgcatga tgaatctgca gggagaggag tttgtgtgcc tcaaatttat tattttgttt 1860
 aattctggag tgtacacatt tctgtccagc accctgaatg ctctggaaaga gaaggaccat 1920
 atccaccggc tcctggacaa gatcacagac actttgatcc acctgtggc caaggcagc 1980
 ctgaccctgc agcagcagca ccagggctg gcccagctcc tcctcatctt ctccccacatc 2040
 aggacatgaa gtaacaaagg catggagcat ctgtatagca tgaagtgc aaacgtggg 2100
 cccctctatg acctgctgt ggagatgtg gacgcccacc gcctacatgc gcccactagc 2160
 cgtacggccgg ccgacccctt gacgacttc gacctggaca tgctggccgc cgacgcccctg 2220
 gaeacttcg acctggacat gctgcccggc gacccctgg acgacttcga cctggacatg 2280
 ctgcccgggt aactaagtaa gggccgctc gagtctagag gcccgttta aacecgctga 2340
 tcagcctcga ctgtgccttc tagttgcaccc ccatctgttg ttggccctc cccctgtcct 2400
 tccttgaccc tggaaagggtgc cactcccaact gtccttcctt aataaaatga gaaatttgc 2460
 tcgcattgtc tgatgttgt tcattctatt ctgggggggtt ggggtggggca gggacagcaag 2520
 ggggaggatt gggaaagacaa tagcggcat gttggggatg cggtgggctc tatggcttct 2580
 gaggcggaaa gaaccagctg gggctctagg gggatcccc acgcgcctg tagcggcgc 2640
 ttaagcgcgg cgggtgtggt gtttacgcgc agcgtgaccg ctacacttgc cagcgcctta 2700
 ggcggccgtc ctttcgttt cttcccttcc ttctcgcca cgtttagtacg gcacctcgac 2760
 caagctctaa atcggggcat ccctttaggg ttccgattta cgttagtgggcatcgc 2820
 cccaaaaaac ttgatttaggg ttagtggttca ctttttttttccatcgc 2880
 ttccgcctt tgacgttggc gtccacgtc ttaatagtg gactttgtt ccaaacttgc 2940
 acaacactca accctatctc ggtctattt tttgatttat aagggtttt ggggatttgc 3000
 gcctatttgtt taaaaaatga gctgatttaa caaaaattttt acgcgaatta attctgtgaa 3060
 atgtgttca gtttaggtgtt gggaaagtccc caggctcccc aggcaggcag aagtatgc 3120
 agcatgcattc tcaatttagtca agcaaccagg ttttttttttgc cccaggctc cccaggcaggc 3180
 agaagtatgc aaagcatgca tctcaatttgc tcagcaacca tagtcccggc cctaactccg 3240
 cccatcccgcc ccctaactcc gcccagttcc gcccatttcc cggccatgg ctgactaatt 3300
 ttttttattt atgcagaggc cgaggccggc tctgcctctg agctatttca gaagtagtgc 3360
 ggaggctttt ttggaggccct aggcttttgc aaaagctcc cgggagcttgc tatatccatt 3420
 ttccgatctg atcaagagac aggttggatc tcgtttgc tgatgttgc 3480
 cacgcgggtt cttccggccgc ttggggggatc aggttattcg gctatgactg ggcacaacag 3540
 acaatccggc gctctgtatgc cggcgttgc cggctgtcag cgcaggggcg cccgggttctt 3600
 tttgtcaaga ccgacctgtc cgggtcccttgc aatgaacttgc agacgaggc agcgcggc 3660
 tcgtggctgg ccacgacggg ctttgccttgc gcagctgtc tcgacgttgc cactgaagcg 3720
 ggaaggggact ggctgtattt gggcgaatgtt cggggcagg atctcctgtc atctcacctt 3780
 gctctgtcccg agaaaagtatc catcatggct gatgcaatgc ggcggctgc taatgttgc 3840
 ccggctaccc gcccatttgc ccaccaagcg aaacatgcgc tgcggcggc acgtactccg 3900
 atggaaaggccg gtctgttgc tcaggatgtat gtcggacgc agcatcaggg gtcgcgc 3960
 gccgaactgt tcggcaggctt ccaggcgcgc atggccgcgc gggaggatct cgtcggtacc 4020
 catggcgatc ctttgcgttcc gaatatcatg gtggaaaatgc gccgcttttgc tggatttcatc 4080
 gactgtggcc ggctgggtgtt ggcggaccgc ttttttttttgc tagcgttgc taccctgtat 4140
 attgtctgaag agtttggccgg cgaatgggtt gaccgcttcc tgcgttgc ttttttttttgc 4200
 gctcccgattt ccgcggccat ctttgcgttcc ttttttttttgc acgagttttt ctgagcggg 4260
 ctctgggggtt cggaaatgacc gaccaaggca cggccaaacctt gccatcacga gatttcgatt 4320
 ccaccggccgc ctttgcgttcc aggttgggtt tcggaaatgc tttccgggac gccggctgc 4380
 tgatccttccatgc ggcggccggat ctttgcgttcc agtttgcgtt ccaccccaac ttgtttatgc 4440
 cagtttataa tggttacaaa taaagcaata gcatcacaat ttttttttttgc ttgtttatgc 4500
 ttctacttgc ttttttttttgc aacttgcgtt ttttttttttgc ttgtttatgc 4560
 taccgtcgatc ctcttagtgc agtttgggtt aatcttgcgtt atagctgtttt cctgtgtgaa 4620
 atttttttttgc gctcacaattt ccacacaacaa tacggccgg aagcataaag tgtaaagctt 4680
 ggggtgccta atggatgttgc taacttgcgtt taatttgcgtt ggcgttactg cccgtttcc 4740
 agtcggggaaa ctttgcgttcc cagttgcgtt aatgttgcgtt ccaacgcgcg gggagaggcg 4800
 gtttgcggcg agcggatataa gctcacttgc aggggttttgc acggttattcc acagaatcag 4860
 gggataacgc agggaaagaac atgttgcgtt aatgttgcgtt aaggccagg aaccgtaaaa 4920
 aggccgcgtt gcttgcgtt ttccataggc tccggccccc tgacgagcat cacaatatac 4980
 gacgctcaag tcagaggtgg cggaaaccgc caggactata aagataccag gcgtttcccc 5040

ctggaaagtc	cctcgtgcgc	tctccgttgc	cgaccctgcc	gcttaccggc	tacctgtccg	5100
cctttctccc	ttcgggaagc	gtggcgctt	ctcaatgctc	acgctgttagg	tatctcagt	5160
cggtgttaggt	cgttcgtc	aagctggct	gtgtgcacga	acccccgtt	cagcccgacc	5220
gctgcgcctt	atccggtaac	tatcgcttgc	agtccaaaccc	ggtaagacac	gacttatcgc	5280
cactggcagc	agccactgg	aacaggatt	gcagagcgag	gtatgttaggc	ggtgctacag	5340
agttcttggaa	gtgggtggct	aactacgg	acactagaag	gacagtattt	ggtatctcgc	5400
ctctgtctaa	gcca	ttcgaaaaaa	gagttggtag	ctcttgatcc	ggcaaaca	5460
ccaccgtctgg	tagcgggtgg	ttttttgttt	gcaagcagca	gattacgcgc	agaaaaaaaaa	5520
gatctcaaga	agatccttgc	atctttctca	cggggtctga	cgctcagtgg	aacgaaaact	5580
cacgtaagg	gat	atgagattat	caaaaaggat	cttcacccat	atccttttaa	5640
attaaaaatg	aagt	tcaatctaaa	gtatataatga	gtaaaacttgg	tctgacagtt	5700
accaatgctt	aatcagt	gcacccatct	cagcgatctg	tctat	tcatccatag	5760
ttgcctgact	ccccgtcg	tagataacta	cgatacggg	gggcttacca	tctggcccca	5820
gtgctgaat	gataccgcga	gaccacgc	caccggctcc	agatttatca	gcaataaaacc	5880
agccagccgg	aagg	cgccagaatg	gtctcgtcaac	tttacccgc	tccatccagt	5940
cttataattg	ttgc	gttagat	gtatgtcc	agttaatagt	ttgcgcac	6000
ttgttgc	tgt	acaggc	atcggtgt	cacgctcg	gttggat	6060
gctccgggtt	ccaa	cgatca	aggcgagtt	catgatcccc	catgttgc	6120
ttagctcc	cgg	cgttgc	atcggttca	gaagtaagtt	ggccgcagt	6180
tggttatggc	agca	actgc	aattcttca	ctgtcat	atccgtaa	6240
tgactgggt	gtact	caacc	aagtcat	gagaatag	tatgcggcga	6300
cttgc	gtca	ataacgg	gataatacc	cgccacat	cagaacttta	6360
tcattggaaa	acgtt	tctcg	ggcgaaaac	tctcaaggat	cttaccgc	6420
ttcgtatgt	acc	actcg	gcaccaact	gatcttc	atctttact	6480
tttctgggt	agca	aaaaaaca	ggaaggc	atccgc	ttcaccagcg	6540
ggaaatgtt	aata	cata	cttcc	ttcaatatta	ttgaagcatt	6600
attgtctcat	gagcggat	ac	atattgt	gtat	tatcagggtt	6660
cgcgacatt	tcc	ccgaaaaa	gtgc	acgtc		6695

```
<210> 15
<211> 6695
<212> DNA
<213> Artificial Sequence
```

<220>

<220>
<223> Description of Artificial Sequence: Construct
LBDBSG521R

```

<400> 15
gacggatcg  gagatctccc gatcccctat ggtcgactct cagtacaatc tgctctgatg 60
ccgcatagtt aagccagtt ctgctccctg ctttgtgtt ggaggtcgct gagtagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatt aagaatctgc 180
tttagggttag gcgccccggc ctgttcgcg atgtacggc cagatatacg cggttgcatt 240
gattattgac tagttataa tagtaatcaa ttacggggtc attagttcat agccccatata 300
ttggatcccg cgttacataa cttacggtaa agggccgc tggctgaccg cccaacgacc 360
cccgcccccatt gacgtcaata atgacgtatg ttccccatagt aacgccaata gggactttc 420
attgacgtca atgggtggac tatttacggt aaactgccc cttggcagta catcaagtgt 480
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaaatggccc gcctggcatt 540
atgcccagta catgaccta tgggactttc ctacttgca gtacatctac gtattagtca 600
tcgctattac catggtgatg cggtttggc agtacatcaa tgggcgtgga tagcggttt 660
actcacgggg atttccaagt ctccacccca ttgacgtcaa tggaggtttg ttttggcacc 720
aaaatcaacg ggactttcca aaatgtcgt acaactcgc cccattgacg caaatggcg 780
gttaggcgtgt acgggtggag gtctatataa gcagagctt ctggctaact agagaaccca 840
ctgtcttactg gcttatcgaa attaatacgca ctcaactatag ggagacccaa gctggctagc 900
gtttaaactt aagcttggta ccgagctcg atccactatgt ccagtgtggt ggaattcctg 960
cagccccgggg gatctatggc ccaggcggcc ctcgagccct atgcttgcct tgctcgatcc 1020
tgcgatcgcc gctttctaa gtcggctgat ctgaagcgcc atatccgcat ccacacaggc 1080
cagaagcctt tccagtgtcg aatatgcatt cgttaacttca gtctgtatgtc ccacacctacc 1140
acccacatcc gcacccacac aggcgagaag cctttgcct gtgacatttg tgggaggaag 1200
tttgcctggaa gtgtatgaacg caagaggcat accaaaatcc attaagacaa gagggactct 1260
agaactatgt accgaagagg agggaatgt ttgaaacaca aggcggcagag agatgtatggg 1320
gagggcaggg gtgaagtggg gtctgttggc gacatggagag ctggccaaacct ttggccaaagc 1380
ccgctcatga tcaaacgcctc taagaagaacg agcctggct tggccctgtac ggccgaccag 1440
atggtcgtgt ccttggta tgctgagccc cccatactct attccgagta tgatcttacc 1500
agacccttca gtgaagcttc gatgtatggc ttactgacca acctggcaga cagggagctq 1560

```

gttcacatga	tcaactgggc	gaagagggtg	ccaggcttg	tggatttgc	cctccatgat	1620
caggtccacc	ttctagaatg	tgcctggcta	gagatcctga	tgattggct	cgtctggcgc	1680
tccatggagc	acccaggaa	gctactgttt	gctcttaact	tgctcttgg	caggaaccag	1740
ggaaaatgtg	tagagggcat	ggtggagatc	ttcgacatgc	tgctggctac	atcatctcg	1800
ttccgcatga	tgaatctgca	gggagagggag	tttgtgtgcc	tcaaatttat	tatttgctt	1860
aattctggag	tgtacacatt	tctgtccagc	accctgaagt	ctctggaaaga	gaaggaccat	1920
atccaccgag	tcctggacaa	gatcacagac	actttgatcc	acctgatggc	caaggcaggc	1980
ctgaccctgc	agcagcagca	ccagcggctg	gccagctc	tcctcatcct	ctcccacatc	2040
aggcacatgc	gtaacaaacg	catggagatc	ctgtacagca	tgaagtcaa	gaacgtggtg	2100
ccccctctatg	acctgctgct	ggagatgctg	gacccccacc	gcctacatgc	gcccactagc	2160
cgtacgcccgg	ccgacccct	ggacgacttc	gacctggaca	tgctggccgg	cgacgcccctg	2220
gacgacttcg	acctggacat	gctgccggcc	gacgcccctgg	acgacttcga	cctggacatg	2280
ctgcccgggt	aactaagtaa	gcccccgctc	gagtttagag	ggcccggtta	aaccgcgtga	2340
tcagcctcga	ctgtgccttc	tagttgccag	ccatctgtt	tttgcccttc	ccccgtgct	2400
tccttgaccc	tggaaagggtc	cactccact	gtccttccct	aataaaatga	gaaattgca	2460
tcgcattgtc	tgagtaggtg	tcattctatt	ctgggggggtg	gggtggggca	ggacagcaag	2520
ggggaggatt	ggaaagacaa	tagcaggatc	gctggggatg	cggtggtct	tatggcttct	2580
gaggcggaa	gaaccagctg	gggtctttagg	gggtatcccc	acgcggcttgc	tagcggcgc	2640
ttaagcgg	cgggtgttgt	ggttacgcgc	agctgtgaccg	ctacacttgc	cagcggcccta	2700
gcccggctc	cttcgcctt	cttcccttcc	tttctcgcca	cgttcgccgg	ctttccccgt	2760
caagctctaa	atcggggcat	ccctttaggg	ttccgattta	gtgcttacg	gcacctcgac	2820
ccccaaaaaa	ttgatttaggg	tgatggttca	cgtagtggc	catcgccctg	atagacggtt	2880
tttcgcctt	tgacgttgg	gtccacgttc	ttaaatagtg	gactcttgc	ccaaactgga	2940
acaacactca	accctatctc	ggtctattct	tttgatttat	aagggatttt	ggggatttcg	3000
gcctattgg	taaaaaatga	gctgatttaa	caaaaattta	acgcgaaatta	attctgtgga	3060
atgtgtgtca	gttaggtgt	ggaaagtccc	cagctcccc	aggcaggcag	aagtatgcaa	3120
agcatgcata	tcaatagtc	accaaccagg	tgtggaaagt	ccccaggctc	cccagcaggc	3180
agaagtatgc	aaagcatgca	tctcaattag	tcagcaacca	tagtcccgcc	cctaactccg	3240
cccatccgc	ccctaactcc	gcccagttcc	gccattctc	cgccccatgg	ctgactaatt	3300
tttttattt	atgcagaggc	cgaggccggc	tctgcctctg	agctattcca	gaagtagtga	3360
ggaggctttt	ttggaggcct	aggctttgc	aaaaagctcc	cgggagctt	tatatccatt	3420
ttcggatctg	atcaagagac	aggatgagga	tcgtttcgca	tgattgaaca	agatggattg	3480
cacgcagggt	ctccggccgc	ttgggtggag	aggctattcg	gctatgactg	ggcacaacag	3540
acaatccgc	gctctgtatgc	cgccgtgttc	cggctgtcag	cgcaggggcgc	cccggttctt	3600
tttgtcaaga	ccgacctgtc	cggtggccctg	aatgaactgc	aggcaggcgc	agcgcggctc	3660
tcgtggctgg	ccacgacggg	cgttccttgc	gcagctgtgc	tcgacgttgc	cactgaagcg	3720
ggaagggact	ggctgtattt	gggcaagtg	ccggggcagg	atctctgtc	atctcacctt	3780
gctcctgcgg	agaaaagtatc	catcatggct	gatgcaatgc	ggcggctgca	tafcgttgc	3840
ccggctaccc	gcccatttcga	ccaccaagcg	aaacatcgca	tcgagcgagc	acgtactcgg	3900
atggaagccg	gtcttgcga	tcaggatgt	ctgacgcaag	agcatcaggg	gctcgccca	3960
gccgaactgt	tcgccaaggct	caaggcgcgc	atgcccgcacg	gchgaggatct	cgtcgtgacc	4020
catggcgatg	cctgccttgc	gaatatcatg	gtggaaaatg	gccgcgtttt	tgatttcattc	4080
gactgtggcc	ggctgggtgt	ggcggaccgc	tatcaggaca	tagcgttgc	tacccgtgt	4140
attgtgtaa	agcttggcgg	ogaatgggt	gaccgttcc	tcgtgccttta	cggtatcgcc	4200
gctcccgatt	cgcagcgcatt	cgcccttctat	cgcccttcttgc	acgagttctt	ctgagcggga	4260
ctctgggtt	cgaaatgacc	gaccaagcga	cgcccaacct	gccatcacga	gatttcgatt	4320
ccaccggccgc	cttctatgaa	aggttgggt	tcggaatcgt	tttccggac	gccggctgga	4380
tgatcctcca	gcmcggggat	ctcatgctgg	agttcttcgc	ccaccccaac	ttgtttattt	4440
cagcttataa	tggttacaaa	taaagcaata	gcatcacaaa	tttcacaaat	aaagcatttt	4500
tttcactgca	ttcttagttt	ggtttgtcca	aactcatcaa	tgtatctt	catgtctgt	4560
taccgtcgac	ctctagctag	agcttggct	aatcatggtc	atagctttt	cctgtgtgaa	4620
attgttatcc	gctcacaatt	ccacacaaca	tacgagccgg	aagcataaaag	tgtaaagcct	4680
gggggtgccta	ataggatgac	taactcacat	taattgcgtt	gcgctcactg	cccgctttcc	4740
agtctggaaa	cctgtcgatc	cagctgcatt	aatgaatcgg	ccaacgcgcg	gggagaggcg	4800
gtttcgccgc	agcggatata	gctcactcaa	aggcgtaat	acggttatcc	acagaatcag	4860
gggataaacgc	agggaaaagaa	atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	4920
aggccgcgtt	gctggcgttt	ttccataggc	tccgcggcccc	tgacgagcat	cacaaaaatc	4980
gacgctcaag	tcagaggtgg	cgaaaccgga	cagactata	aagataccag	gcgtttcccc	5040
ctggaagctc	cctcgatgc	tctccctgttc	cgaccctgc	gcttaccgga	tacctgtcgg	5100
cctttctccc	ttcgggaagc	gtggcgctt	ctcaatgc	acgctgttagg	tatctcagtt	5160
cggtgttaggt	cgttcgcgtcc	aagctgggt	gtgtgcacga	accccccgtt	cagcccgacc	5220
gctgcgcctt	atccgtaatc	tatcgtcttgc	agtccaaatcc	ggtaaagacac	gacttatcgc	5280
cactggcaggc	agccacttgtt	aacaggat	gcagagcgcag	gtatgttaggc	ggtgtctacag	5340
agttcttgaa	gtgggtggct	aactacggct	acactagaag	gacagtatttt	gttatctgcg	5400
ctctgctgaa	gccagttacc	ttcggaaaaaa	gagttggtag	ctcttgcattcc	ggcaaaacaaa	5460
ccaccgctgg	tagcgggtgg	tttttgc	gcaagcagca	gattacgcgc	agaaaaaaaag	5520
gatctcaaga	agatccttgc	atctttctta	cggggtctga	cgctcagtgg	aacgaaaact	5580

cacgttaagg	gattttggc	atgagattat	caaaaaggat	cttcacccat	atccctttaa	5640
ataaaaatg	aagtttaaa	tcaatctaaa	gtatatatga	gtaaacctgg	tctgacagt	5700
accaatgctt	aatcagttag	gcacctatct	cagcgatctg	tctatttcgt	tcatccatag	5760
ttgcctgact	ccccgtcgt	tagataacta	cgatacggga	gggcttacca	tctggcccc	5820
gtgtgcaat	gataccgcga	gacccacgct	caccggctcc	agatttatca	gcaataaacc	5880
agccagccgg	aaggggcgag	cgcagaagt	gtcctgcaac	tttattccgc	tccatccagt	5940
ctattaattg	ttgcccggaa	gctagagtaa	gtagttcgcc	agttaatagt	ttgcgcaacg	6000
tttgttgcct	tgctacaggc	atcggttgt	cacgctcg	gttggatg	gttcattca	6060
gctccgggtc	ccaaacgatca	aggcagtagt	catgatcccc	catgttgc	aaaaaaagcgg	6120
tttagctctt	cggccctccg	atcggtgtca	gaagtaagtt	ggccgcagt	ttatcactca	6180
tgggtatggc	agcaactgcat	aattctctta	ctgtcatg	atccgtaaga	tgcctttctg	6240
tgactggtga	gtactcaacc	aagtcaattt	gagaatagt	tatgcggcga	ccgagttgt	6300
cttgcggcgc	gtcaatacgg	gataataccg	cgccacatag	cagaacttta	aaagtgc	6360
tcattggaaa	acgttctcg	ggcgaaaaac	tctcaaggat	cttaccgctg	ttgagatcca	6420
gttcgatgt	acccaactcg	gcacccaaact	gatcttcagc	atctttact	ttcaccagcg	6480
tttctgggtg	agcaaaaaca	ggaaggcaaa	atgcgc	aaaggaaata	aggcgcacac	6540
ggaatgtt	aatactcata	ctttcctt	ttcaatatta	ttgaagcatt	tatcagggtt	6600
attgtctcat	gagcggatac	atatttgaat	gtatattagaa	aaataaaca	ataggggttc	6660
cgccacatt	tccccgaaaa	gtgccac	acgtc			6695

<210> 16
<211> 6801
<212> DNA
<213> Artificial Sequence

<220>

<220>
<223> Description of Artificial Sequence: Construct
LBDBSVP16

<400> 16						
gacggatcg	gagatctccc	gatcccstat	ggtcgactct	cagtacaatc	tgctctgat	60
ccgcatagtt	aagccagt	ctgccccctg	cttgggtgtt	ggaggtcg	gagtagtgc	120
cgagcaaaat	ttaagctaca	acaaggcaag	gcttgcac	caattgc	aagaatctgc	180
ttagggttag	gcttttgcg	ctgttcgcg	atgtacgggc	cagatatacg	cgttgcac	240
gattattgac	tagttattaa	tagtaatcaa	ttacgggtc	attagttcat	agccatata	300
tggagttccg	cgttacataa	cttacggtaa	atggccgc	tggctgacc	cccaacgacc	360
ccgcgcatt	gacgtcaata	atgacgtat	ttccatagt	aacgccaata	gggacttcc	420
attgacgtca	atgggtggac	tatttacgt	aaactgccc	cttggcagta	catcaagt	480
atcatatgcc	aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	540
atgcccagta	catgaccta	tgggactt	ctacttggca	gtacatctac	gtattagtc	600
tcgctattac	catggtgat	cggttttgc	agtacatcaa	tggcgtg	tagcggttt	660
actcacgggg	atttccaat	ctccacccca	ttgacgt	tggaggtt	tttggcacc	720
aaaatcaacg	ggacttcca	aaatgtcgt	acaactccgc	ccattgacg	caaatggcg	780
gtagggcgt	acgggtggag	gtctatataa	gcagagct	ctggctaact	agagaaccca	840
ctgcttactg	gcttatcgaa	attaatacga	ctcactatag	ggagaccaa	gctggctag	900
gtttaaactt	aagcttggta	ccgagctcg	atccactagt	ccagtgtg	ggaattcct	960
cagcccgggg	gatctatggc	ccaggcggc	ctcgagcc	atgcttccc	tgtcgagt	1020
tgcgatcg	gttttctaa	gtcggtgt	ctgaagcgc	atatccgc	ccacacaggc	1080
cagaaggcctt	tccagtgtcg	aatatgc	ctgtaacttca	gtcgtagt	ccaccttacc	1140
accacatcc	gcacccacac	aggcgagaag	cctttgc	gtgacattt	tgggaggaag	1200
tttgcagga	gtgtgaac	caagaggcat	acaaaatcc	atthaagaca	gaggact	1260
agaactagt	accgaaggagg	agggagaat	ttgaaacaca	agcgcacag	agatgatgg	1320
gagggcagg	gtgaagtgg	gtctgctg	gacatgagag	ctgccaac	ttggccaagc	1380
ccgctcatg	tcaaactc	taagaagaa	agcctggc	tgtccctg	ggccgacc	1440
atggtcgt	cttgttgg	tgctgagcc	cccatact	atccgagta	tatcctacc	1500
agacccttca	gtgaagctc	gatgtggc	ttactgacca	acctggcaga	cagggagct	1560
gttcacatg	tcaactggc	gaagagggt	ccaggtt	tggattgac	cctccatgat	1620
caggtccacc	ttctagaat	tgcctggc	gagatc	tgattgg	cgtctggc	1680
tccatggagc	acccagg	gctactgtt	gctctact	tgccttg	caggaacc	1740
ggaaaatgt	tagggc	gttggagatc	ttcgacat	tgcgttac	atcatctcg	1800
ttccgcata	tgaatcgt	ggggagagg	tttgg	tcaaatctat	tat	1860
aattctggag	tgtacacatt	tctgtcc	accctga	ctctggaa	gaaggaccat	1920
atccaccgg	tcctggacaa	gatcacagac	acttgc	acctgtat	caaggcaggc	1980
ctgaccctgc	agcagcagca	ccagcggct	gcc	tcctcatc	ctcccacatc	2040
aggcacatg	gtaacaaagg	catggagat	ctgtac	tgaagt	gaacgtgg	2100

ccccctatcg	acctgctgt	ggagatgctg	gacgcccacc	gcctacatgc	gcccaactgc	2160
cgtacgggct	ctcccccgtc	cgatgtcagc	ctggggacg	agctccactt	agacggcag	2220
gacgtggcgt	tggcgtatgc	cgacgcgcta	gacgattcg	atctggacat	gttggggac	2280
ggggattttcc	cgggtccggg	atttaccccc	cacgactccg	ccccctacgg	cgctctggat	2340
atggccgact	tcgagtttg	gcagatgttt	accgatgccc	ttggaaattga	cgagtagcgt	2400
tagggggcg	ccgcgtcgagt	ctagaggggc	cgtttaacc	cgctgtatcg	ctctgactgt	2460
gcctttagt	tgccagccat	ctgtgttttg	ccccctcccc	gtgccttcct	tgacccttgg	2520
agggtccact	cccactgttc	tttcttaata	aatatggggaa	attgcatcgc	attgtcttag	2580
taggtgtcat	tctatttctgg	gggggtgggg	ggggcaggac	agcaaggggg	aggattggga	2640
agacaatagc	aggcatgtcg	gggatgcgg	gggctctatg	gcttctgagg	cggaaaagaac	2700
cagctggggc	tctagggggt	atccccacgc	gccctgttagc	ggcgattaa	gcccggcggg	2760
tgtgggtgtt	acgcgcagcg	tgaccgtac	acttgccagc	gccctagcgc	ccgcctcctt	2820
cgccttcttc	ccttccttgc	tcgcccacgtt	cgccggcttt	ccccgtcaag	ctctaaatcg	2880
gggcattccct	ttagggttcc	gatttagtgc	tttacggcac	ctcgacccca	aaaaactgt	2940
ttaggggtat	ggtcacgt	gtgggccatc	gccctgatag	acgggttttc	gccctttgac	3000
gttggagtcc	acgttcttta	atagtggact	cttggcccaa	atctggacaa	cactcaaccc	3060
tatctcggt	tattcttttg	atttataagg	gattttgggg	atttcggcct	attggttaaa	3120
aaatagatgt	attaacaaa	aatttaacgc	gaatttaattc	tgtgaaatgt	gtgtcagtt	3180
gggtgtggaa	agtccccagg	ctccccaggg	aggcagaagt	atgcaagaca	tgcacatctaa	3240
tttagtcagca	accaggtgt	gaaagtcccc	aggctcccca	gcagggcagaa	gtatgcaaaag	3300
catgcatctc	aattagtca	caaccatagt	cccgccctta	actccgcccc	tcccgcccc	3360
aactccgccc	agttccgccc	atttccgccc	ccatggctga	ctaatttttt	ttatttatgc	3420
agaggccgag	ggccgcctcg	cctotgagct	attccagaag	tagtgaggag	gcttttttgg	3480
aggcctaggc	ttttgcacaa	agctcccggg	agcttgcata	tccattttcg	gatctgtatca	3540
agagacagga	tgaggatcgt	ttcgcatgt	tgaacaagat	ggattgcacg	cagggtctcc	3600
ggccgcttgg	gtggagaggg	tattcggtca	tgactggca	caacagacaa	tcggctgtc	3660
tgatgcggcc	gtgttccggc	tgteagcgc	ggggcggccg	gttcttttgc	tcaagaccga	3720
cctgtccgg	gcccctgaatg	aactcgagg	cgaggcagcg	cggctatcgt	ggctggccac	3780
gacgggggtt	ccttgcgcag	ctgtgtctga	cgttgtca	gaagcggggaa	gggactggct	3840
gctattggc	gaagtgcggg	ggcaggatct	cctgtcatct	caccttgctc	ctgcccggaa	3900
agtatccatc	atggctgtat	caatgcggcg	gctgcatacg	tttgatccgg	ctacactgccc	3960
attcgaccac	caagcgaaac	atcgcatcg	gcgagcacgt	actcggatgg	aagccggct	4020
tgtcgatcag	gatgtatctgg	acgaagagca	tcaggggctc	ggcccgccgg	aacttgtcgc	4080
caggctcaag	ggcgcatgtc	ccgacggcga	ggatctcg	gtgaccatcg	gcgtatcctg	4140
cttggccat	atcatgggtt	aaaatggccg	ctttcttgg	ttcatcgact	gtggccggct	4200
gggtgtggcg	gaccgtatac	aggacatagc	gttggctacc	cgtgatattg	ctgaagagat	4260
tggcggcgaa	tgggctgacc	gcttctcg	gctttacgg	atcggccgtc	ccgattcgca	4320
gogcatogcc	ttctatcgcc	ttcttgacga	gttcttctga	gccccgactct	ggggttcgaa	4380
atgaccgacc	aagcgacgccc	caacctgcca	tcacgagatt	tcgattccac	cgccgccttc	4440
atgaaaagg	tgggcttccgg	aatgttttc	cgggaccccg	cggtggatgt	cctccagcgc	4500
ggggatctca	tgctggaggt	cttcgcccac	ccaaacttgc	ttatttgacg	ttataatgg	4560
tacaaataaa	gcaatagcat	cacaatttc	acaaataaaag	catttttttc	actgcattct	4620
agttgtggt	tgccaaact	catcaatgt	tcttatcatg	tctgtatacc	gtcgacctt	4680
agctagatgt	tgccgtatac	atggtcatag	ctgtttctgt	tgtgaaattg	ttatccgctc	4740
acaattccac	aaacatacag	agccggaaagc	ataaaatgt	aaagctgggg	tgcctaatga	4800
gtgagctaac	tcacattaaat	tgcgttgcgc	tcactgccc	cttccagtc	gggaaacactg	4860
tcgtgcggc	tgcatatgt	aatcgccaa	cgcgccggga	gaggcgggtt	gccccggagcg	4920
gtatcagctc	actcaaaggc	gtaatacgg	ttatccacag	aatcagggg	taacgcagga	4980
aagaacatgt	gagcaaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaggc	cgcgttgctg	5040
gcgttttcc	ataggctcg	ccccctgtac	gagcatcaca	aaaatcgacg	ctcaagtctg	5100
aggtgtggc	acccgacagg	actataaaga	taccagggt	ttccccctgg	aagctccctc	5160
gtgcgtctc	ctgttccgac	ctgtccgctt	acggatacc	tgtcggcc	tctcccttcg	5220
ggaagcggtg	cgctttctca	atgtcgtacgc	tgttaggtatc	tcgttccgt	gtaggtcg	5280
cgctccaa	tgggtgtgt	gcacaaaccc	cccggttcc	ccgaccgcgt	cgccttaccc	5340
ggttaactatc	gtttagtgc	caacccggta	agacacgact	tatcgccact	ggcagcagcc	5400
actggtaaca	ggatttagcag	agcgaggtat	gttaggcgtg	ctacagagtt	cttgaagtg	5460
tggcttaact	acggctacac	tagaaggaca	gtatttggta	tctgcgtct	gctgaagcca	5520
gttacccctcg	gaaaaagagt	tggtagctt	tgatccggca	aacaaaccac	cgctggtagc	5580
ggtggtttt	ttgtttgcaa	gcagcagatt	acgcgcagaa	aaaaaggatc	tcaagaagat	5640
ccttgcattt	tttctacggg	gtctgacgt	cagtggaaacg	aaaactcact	ttaagggtt	5700
ttgggtcatga	gattatcaa	aaggatcttc	acctagatcc	ttttaaattt	aaaatgaagt	5760
tttaaatcaa	tctaaagtat	atatgagtaa	acttggctcg	acagtttacca	atgtcttaatc	5820
agtggggcac	ctatctcagc	gatctgtcta	tttcgttcat	coatagttgc	ctgaetcccc	5880
gtcgtgtaga	taactacgt	acgggagggc	ttaccatctg	gccccagtc	tgcaatgata	5940
ccgcgagacc	cacgcctacc	ggctccagat	ttatcagca	taaaccagcc	agccggaaagg	6000*
gcccggcgca	gaagtggtc	tgcaacttta	tccgcctcca	tccagtctat	taattgttgc	6060
cqqqaqacta	qaqtaaqtag	ttcgccagtt	aatagttgc	gcaacgttgc	tgccatttgct	6120

```

acaggcatcg tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttccaa 6180
cgatcaaggc gagttacatg atccccatg ttgtgaaaaa aagcggttag ctccctcggt 6240
cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatgg tatggcagca 6300
ctgcataatt ctcttactgt catgcccattc gtaagatgt tttctgtac tggtagtac 6360
tcaaccaagt cattctgaga atagtgtatc cggcgaccga gttgcttgc cccggcgta 6420
atacggata ataccgcgc acatagcaga actttaaaag tgctcatcat tggaaaacgt 6480
tcttcgggca gaaaactctc aaggatctt ccgcgttga gatccagttc gatgtAACCC 6540
actcgtgcac ccaactgatc ttcagcatct tttactttca ccagcggttc tgggtgagca 6600
aaaacaggaa ggcaaaatgc cgcaaaaaaag ggaataaggg cgacacggaa atgttgaata 6660
ctcatactct tccttttca attattattga agcatttatac agggttatttgc tctcatgagc 6720
ggatacatat ttgaatgtat tttagaaaaat aaacaaatag gggttccgcg cacatttccc 6780
cgaaaagtgc cacctgacgt c 6801

```

<210> 17
<211> 1551
<212> DNA
<213> Artificial Sequence

<220>

<220>
<223> Description of Artificial Sequence: Construct
VP16C7ER

<400> 17

```

gctagcgcca ccatggggcg cgccggcgct ccccccggccg atgtcagcct gggggacgag 60
ctccacttag acggcgagga cgtggcgatg ggcgcattccg acgcgcgtaga cgatttcgt 120
ctggacatgt tggggacgg ggattccccg ggtccggat ttacccccc cgaactccgccc 180
ccctacggcg ctctggatat ggccgacttc gagtttgagc agatgtttac cgatgccctt 240
ggaatttgacg agtacggttt attaacaag cttggggccc aggcggccct cgagccctat 300
gctggccctg tcgagtcctg cgatcgccgc ttttctaagt cggctgatct gaagcgccat 360
atccgcattcc acacaggccca gaagcccttc cagttcgaa tatgcattcg taacttcagt 420
cgtagtgacc accttaccac ccacatccgc acccacacag gcgagaagcc tttgcctgt 480
gacattttgtt ggaggaagtt tgccaggagt gatgaacgcg agaggcatac caaaatccat 540
ttaagacaga aggactctag aactagtggc caggccggcc agggggatcc acgaaatgaa 600
atgggtgctt caggagacat gagggtcgcc aacctttggc caagccctct tttgtattaag 660
cacactaaga agaatagccc tgccttgcgtt tgacagctg accagatggt cagtccttg 720
ttggatgtcg aaccggccat gatctatttc gaatatgtat cttctagacc cttcagtgaa 780
gcctcaatga tgggttttattt gaccaaccta gcagatggg agctgggttc tatgtacac 840
tggcggaaaga gagtggccagg ctttggggac ttgaaatctcc atgatcaggat ccaccttctc 900
gagttgtccct ggctggagat tctgtatggat ggttcgtctt ggcgctccat ggaacacccg 960
ggaaagctcc tgggtgtcc taacttgcgtt ctggacagga atcaaggtaa atgtgtggaa 1020
ggcatgggtgg agatcttga catgttgcgtt gctacgtcaa gtcgggtccg catgtgaac 1080
ctgcagggtg aagagtttgcgtt gtgcctcaaa tccatcattt tgcttaattc cggagtgtac 1140
acgtttctgt ccagcacctt gaagtctctg gaagagaagg accacatcca ccgtgtccctg 1200
gacaagatca cagacacttt gatccacccgtt atggccaaag ctggccctgac tctgcagcag 1260
cagcatcgcc gcctagctca gtccttctc attctttccc atatccggca catgagtaac 1320
aaaggcatgg agcatctca caacatgaaa tgcaagaacgc ttgtgcctt ctatgacctg 1380
ctcctggaga tgggtggatgc ccacccgcattt catgcccccg ccagtcgttgc gggagtgc 1440
ccagaggagc ccagccagac ccagtcggcc accaccagat ccacttcagc acattcctta 1500
caaacctact acataccccc ggaaggcagag ggcttccca acacgatctg a 1551

```

<210> 18
<211> 1404
<212> DNA
<213> Artificial Sequence

<220>

<220>
<223> Description of Artificial Sequence: Construct
VP16C7PR

<400> 18

```

gctagcgcca ccatggggcg cgccggcgct ccccccggccg atgtcagcct gggggacgag 60
ctccacttag acggcgagga cgtggcgatg ggcgcattccg acgcgcgtaga cgatttcgt 120
ctggacatgt tggggacgg ggattccccg ggtccggat ttacccccc cgaactccgccc 180

```

ccctacggcg ctctggatat ggccgacttc gagtttgagc agatgttac cgatgccctt 240
 ggaattgacg agtacggttt attaacaag cttggggccc aggccgcctt cgagccctat 300
 gcttgcctg tcgagtccctg cgatcgccgc ttttctaagt cggctgatct gaagcgccat 360
 atccgcattcc acacaggcca gaagcccttc cagtgtcgaa tatgcattcg taacttcagt 420
 cgttaacttaccac ccacatccgc acccacacag gcgagaagcc ttggcctgt 480
 gacatttgcg ggaggaagtt tgccaggagt gatgaacgca agaggcatac caaaatccat 540
 ttaagacaga aggactctag aactagtggc caggccggcc aggggatcc agtcagagg 600
 gtgagagcac tggatgtctgt tgctctccca cagccagtgg gcgttccaaa tgaaagccaa 660
 gcccataagcc agagattcac ttttccacca ggtcaagaca tacagtttat tccaccactg 720
 atcaacactgt taatgagcat tgaaccagat gtgatctatg caggacatga caacacaaaa 780
 cctgacaccc ccagttctt gctgacaagt cttaatcaac taggcgagag gcaacttctt 840
 tcagtagtca agtggctcaa atcattgcca ggtttcgaa acttacatat ttagtgcac 900
 ataactctca ttcagtttc ttggatgagc ttaatggtgt ttggcttagg atggagatcc 960
 tacaaacacg tcagtggca gatgctgtat tttgcacccgc atctaatact aaatgaacacag 1020
 cggatgaaag aatcatcattt ctattcatat tgcccttacca tggcagat cccacaggag 1080
 ttgtcaagc ttcaagtttag ccaagaagag ttccctctgtat gaaagtatt gttacttctt 1140
 aatacaattc ctttggaaagg gctacgaagt caaacccagt ttgaggagat gaggtcaagc 1200
 tacatttagag agtcatcaa ggcaatttgtt tgaggcataa aaggaggatgt gtcgagctca 1260
 cagcgtttct atcaacttac aaaacttctt gataacttgc atgatcttgc caaacaactt 1320
 catctgtact gcttgaatac atttatccag tcccgccac tgagtgttga atttccagaa 1380
 atgatgtctg aagttattgc ttga 1404

<210> 19
 <211> 5
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Recombinant molecule

<400> 19
 Thr Gly Glu Lys Pro
 1 5

<210> 20
 <211> 19
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Recombinant molecule
 <220>

<221> n= {N}x ; X= any number

<222> 10

<400> 20
ggcccacgcn gcgtggcg 19

<210> 21
 <211> 28
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Recombinant molecule
 <220>

<221> n= {N}x; X= any number

<222> 19

<400> 21
cgccggccgcgccc gccgccccng cgtggggcgcg 28

<210> 22
<211> 35
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 22
Met Lys Leu Leu Glu Pro Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg
1 5 10 15

Arg Phe Ser Lys Ser Ala Asp Leu Lys Arg His Ile Arg His Thr Gly
20 25 30

Glu Lys Pro
35

<210> 23
<211> 29
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 23
Tyr Ala Cys Pro Val Glu Ser Cys Asp Arg Arg Phe Ser Lys Ser Ala
1 5 10 15

Asp Leu Lys His Ile Arg Ile His Thr Gly Glu Lys Pro
20 25

<210> 24
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 24
cctcgccgcgcgccc gcgggttttc ccgcgcgcgcgccc gagg 34

<210> 25
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<220>

<221> nnn= a mixture of all 64 existing triplets and its complement

<222> 26-28 and 7-9 respectively

<400> 25
ggacgcnnc gcgggtttc cgcgnnnnc gtcc 34

<210> 26
<211> 66
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 26
gcgagcaagg tcgcggcagt cactaaaaga tttgccgcac tctggcatt tatacggtt 60
ttcacc 66

<210> 27
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 27
gtgactgccc cgaccttgct cgcgcataac gcactcatac tggcgagaag ccataacaat 60
gtccagaatg tggc 74

<210> 28
<211> 81
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 28
ggtaagtcc tctctcagag ctctcacctg gtgcgccacc agcgtaccca cacgggtgaa 60
aaaccgtata aatgcccaga g 81

<210> 29
<211> 58
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 29
acgcaccagg ttgtcagagc ggctgaaaga cttgccacat tctggacatt tgtatggc 58

<210> 30
<211> 87
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 30
gaggaggagg aggtggccca ggccggccctc gagcccgaaaa agaagcccta tgcttgtccg 60
gaatgtggta agtccttctc tcagagc 87

<210> 31
<211> 81
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 31
gaggaggagg agctggccgg cctggccact agtttttta ccggtgttag tacgttggtg 60
acgcaccaggc ttgtcagagc g 81

<210> 32
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 32
gaggaggagg ctagcgggat gtggtcttgc cctcaacagg tagg 44

<210> 33
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 33
gaggaggaga agcttctcggt ccgcctcccg cggcgctccg c 41

<210> 34
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 34
gaggaggagg ctagccgatg tgactgtctc ctcccaaatt tgttagacc 48

<210> 35
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 35
gaggaggaga agcttgggtgc tcactgcggc tccggccccca tg 42

<210> 36
<211> 11
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 36
Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu
1 5 10

<210> 37

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 37

gaggagggct gcttgaggaa gta

23

<210> 38

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 38

gccggagcca tggggccgga gcc

23

<210> 39

<211> 48

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 39

cctactgccg gcactagttc tgctggagac atgagagctg ccaacctt

48

<210> 40

<211> 42

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 40

cctaaacgta cggctagtgg ggcgtatgtag gcggtggcgt tc

42

<210> 41

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 41

cctaaacgta cggactgtgg cagggaaacc ctctgcctc

39

<210> 42
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 42
ccacttaaat gtgaaaagtcg tacgccggcc 30

<210> 43
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 43
tatggggggc tcagcatcca acaaggcact 30

<210> 44
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 44
cctactacta gtgaccgaag aggagggaga atgttcaaac acaagcgc 48

<210> 45
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 45
cctactacta gtagtattca aggacataac gactatatgt gt 42

<210> 46
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 46
tatcatgtgc ggcccgttac tttagttaccc cggcagcat 39

<210> 47
<211> 39
<212> PRT
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 47
Pro Ala Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Pro Ala Asp
1 5 10 15
Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Pro Ala Asp Ala Leu Asp
20 25 30
Asp Phe Asp Leu Asp Met Leu
35

<210> 48

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 48

gatccaaagt cgcgtggcg cagcgccac gcgatcaaag a

41

<210> 49

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 49

gatccaaagt ccaggcgagc gcgtggcg cagatcaaag a

41

<210> 50

<211> 47

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 50

gatccaaagt cgcgtggcg caggcgcgag cgtggcgga tcaaaga

47

<210> 51

<211> 41

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Recombinant molecule

<400> 51

gatccaaagt cgcgtggcg cagcgccac gcgatcaaag a

41

<210> 52

<211> 41

<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant
molecule

<400> 52
gatccaaagt cgcgtggcg cactccggcc ccgatcaaag a 41

<210> 53
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant
molecule

<400> 53
gatccaaagt cggggccgga gactccggcc ccgatcaaag a 41

<210> 54
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant
molecule

<400> 54
gccggagcca tggggccgga gcc 23

<210> 55
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant
molecule

<400> 55
cgctccctct caggcgcagg g 21

<210> 56
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant
molecule

<400> 56
ggcgccccact gtggggcggg c 21

<210> 57
<211> 41
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant
molecule

<400> 57
gaggaggagg gccggccggg aagccgtgca ggaggagcgg c 41

<210> 58
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 58
gaggaggagg ggcgcgccag tcatttggtg cggcgctcc agc 43

<210> 59
<211> 42
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 59
gaggaggagt taattaaagt catttggtgc ggccgcctcca gc 42

<210> 60
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 60
gaggaggagg gccggccggg gtggcggcca agactttgtt aagaagg 47

<210> 61
<211> 50
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 61
gaggaggagg gcccaggcgg ccgggtggcgg ccaagacttt gttaagaagg 50

<210> 62
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 62
gaggaggagg ggcgcgccgg catgaacgtc ccagatctcc tcgag 45

<210> 63
<211> 46
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant

molecule

<400> 63
gaggaggagg gccggccgga ggcctgaatg tgtcatacag gagccc 46

<210> 64
<211> 49
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 64
gaggaggagg gcccaggcg ccaggcctga atgtgtcata caggagccc 49

<210> 65
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 65
gaggaggagg gcgcgccccct ccgccacgtc ccagatctcc tcgag 45

<210> 66
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 66
gtacagatgc tccatgcgtt tgttactcat gtgcc 35

<210> 67
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 67
ggcacatgag taacaaacgc atggagcatc tgtac 35

<210> 68
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 68
ccatggagca cccagtgaag ctactgtttg c 31

<210> 69
<211> 31

<212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Recombinant
 molecule
 <400> 69
 gcaaacagta gtttcaactgg gtgctccatg g

31

<210> 70
 <211> 624
 <212> DNA
 <213> Muridae

<220>
 <221> CDS
 <222> (1)...(624)
 <223> cDNA encoding secretion signal and
 murine endostain protein.

<400> 70		
atg gag aca gac aca ctc ctg cta tgg gta ctg ctg ctc tgg gtt cca		48
Met Glu Thr Asp Thr Leu Leu Leu Trp Val Leu Leu Leu Trp Val Pro		
1 5 10 15		
ggt tcc act ggt gac gcg gcc cat act cat cag gac ttt cag cca gtg		96
Gly Ser Thr Gly Asp Ala Ala His Thr His Gln Asp Phe Gln Pro Val		
20 25 30		
ctc cac ctg gtg gca ctg aac acc ccc ctg tct gga ggc atg cgt ggt		144
Leu His Leu Val Ala Leu Asn Thr Pro Leu Ser Gly Gly Met Arg Gly		
35 40 45		
atc cgt gga gca gat ttc cag tgc ttc cag caa gcc cga gcc gtg ggg		192
Ile Arg Gly Ala Asp Phe Gln Cys Phe Gln Gln Ala Arg Ala Val Gly		
50 55 60		
ctg tcg ggc acc ttc cgg gct ttc ctg tcc tct agg ctg cag gat ctc		240
Leu Ser Gly Thr Phe Arg Ala Phe Leu Ser Ser Arg Leu Gln Asp Leu		
65 70 75 80		
tat agc atc gtg cgc cgt gct gac cgg ggg tct gtg ccc atc gtc aac		288
Tyr Ser Ile Val Arg Arg Ala Asp Arg Gly Ser Val Pro Ile Val Asn		
85 90 95		
ctg aag gac gag gtg cta tct ccc agc tgg gac tcc ctg ttt tct ggc		336
Leu Lys Asp Glu Val Leu Ser Pro Ser Trp Asp Ser Leu Phe Ser Gly		
100 105 110		
tcc cag ggt caa gtg caa ccc ggg gcc cgc atc ttt tct ttt gac ggc		384
Ser Gln Gly Gln Val Gln Pro Gly Ala Arg Ile Phe Ser Phe Asp Gly		
115 120 125		
aga gat gtc ctg aga cac cca gcc tgg ccg cag aag agc gta tgg cac		432
Arg Asp Val Leu Arg His Pro Ala Trp Pro Gln Lys Ser Val Trp His		
130 135 140		
ggc tcg gac ccc agt ggg cgg agg ctg atg gag agt tac tgt gag aca		480
Gly Ser Asp Pro Ser Gly Arg Arg Leu Met Glu Ser Tyr Cys Glu Thr		
145 150 155 160		
tgg cga act gaa act act ggg gct aca ggt cag gcc tcc tcc ctg ctg		528
Trp Arg Thr Glu Thr Thr Gly Ala Thr Gly Gln Ala Ser Ser Leu Leu		
165 170 175		
tca ggc agg ctc ctg gaa cag aaa gct gcg agc tgc cac aac agc tac		576

Ser Gly Arg Leu Leu Glu Gln Lys Ala Ala Ser Cys His Asn Ser Tyr
 180 185 190

atc gtc ctg tgc att gag aat agc ttc atg acc tct ttc tcc aaa tag 624
 Ile Val Leu Cys Ile Glu Asn Ser Phe Met Thr Ser Phe Ser Lys *
 195 200 205

<210> 71
 <211> 207
 <212> PRT
 <213> Muridae

<400> 71
 Met Glu Thr Asp Thr Leu Leu Trp Val Leu Leu Trp Val Pro 1 5 10 15
 Gly Ser Thr Gly Asp Ala Ala His Thr His Gln Asp Phe Gln Pro Val 20 25 30
 Leu His Leu Val Ala Leu Asn Thr Pro Leu Ser Gly Gly Met Arg Gly 35 40 45
 Ile Arg Gly Ala Asp Phe Gln Cys Phe Gln Gln Ala Arg Ala Val Gly 50 55 60
 Leu Ser Gly Thr Phe Arg Ala Phe Leu Ser Ser Arg Leu Gln Asp Leu 65 70 75 80
 Tyr Ser Ile Val Arg Arg Ala Asp Arg Gly Ser Val Pro Ile Val Asn 85 90 95
 Leu Lys Asp Glu Val Leu Ser Pro Ser Trp Asp Ser Leu Phe Ser Gly 100 105 110
 Ser Gln Gly Gln Val Gln Pro Gly Ala Arg Ile Phe Ser Phe Asp Gly 115 120 125
 Arg Asp Val Leu Arg His Pro Ala Trp Pro Gln Lys Ser Val Trp His 130 135 140
 Gly Ser Asp Pro Ser Gly Arg Arg Leu Met Glu Ser Tyr Cys Glu Thr 145 150 155 160
 Trp Arg Thr Glu Thr Thr Gly Ala Thr Gly Gln Ala Ser Ser Leu Leu 165 170 175
 Ser Gly Arg Leu Leu Glu Gln Lys Ala Ala Ser Cys His Asn Ser Tyr 180 185 190
 Ile Val Leu Cys Ile Glu Asn Ser Phe Met Thr Ser Phe Ser Lys 195 200 205

<210> 72
 <211> 18
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Integrin β_3 (B3B) target sequence

<400> 72 gcctgagagg gagcggtg 18

<210> 73
 <211> 18
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Integrin β_3 (B3C) target sequence

<400> 73 ggaggggacg cgggtgggt 18

<210> 74
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: ErbB-2 (E2B2) target sequence

<400> 74
gtgtgagaac ggctgcaggc 20

<210> 75
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: ErbB-2 (E2C) target sequence

<400> 75
ggggccggag ccgcagtg 18

<210> 76
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: ErbB-2 (E2D) target sequence

<400> 76
gcagttggag gggcgag 18

<210> 77
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 77
Gln Ser Ser Asn Leu Val Arg
1 5

<210> 78
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 78
Asp Pro Gly Asn Leu Val Arg
1 5

<210> 79
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 79
Arg Ser Asp Asn Leu Val Arg

1 5

<210> 80
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 80
Thr Ser Gly Asn Leu Val Arg
1 5

<210> 81
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 81
Gln Ser Gly Asp Leu Arg Arg
1 5

<210> 82
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 82
Asp Cys Arg Asp Leu Ala Arg
1 5

<210> 83
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 83
Arg Ser Asp Asp Leu Val Lys
1 5

<210> 84
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 84
Thr Ser Gly Glu Leu Val Arg
1 5

<210> 85
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 85
Gln Arg Ala His Leu Glu Arg
1 5

<210> 86
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 86
Asp Pro Gly His Leu Val Arg
1 5

<210> 87
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 87
Arg Ser Asp Lys Leu Val Arg
1 5

<210> 88
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 88
Thr Ser Gly His Leu Val Arg
1 5

<210> 89
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 89
Gln Ser Ser Ser Leu Val Arg
1 5

<210> 90
<211> 7

<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 90
Asp Pro Gly Ala Leu Val Arg
1 5

<210> 91
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 91
Arg Ser Asp Glu Leu Val Arg
1 5

<210> 92
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Recombinant molecule

<400> 92
Thr Ser Gly Ser Leu Val Arg
1 5

INTERNATIONAL SEARCH REPORT

Intern	al Application No
PCT/EP 00/10430	

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C07K14/47 C07K14/72 C12N15/62 C12N15/86 A61K38/17
C07K14/705

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, MEDLINE, BIOSIS, EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>WO 98 18925 A (BAYLOR COLLEGE MEDICINE ;MALLEY BERT O (US); TSAI MING JER (US); T) 7 May 1998 (1998-05-07) cited in the application</p> <p>figures 10,11; examples 36,37</p> <p>---</p> <p>-/-</p>	<p>1-6, 9, 11-21, 23,24, 26-45, 47-62, 65,67-73</p>

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

° Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

26 January 2001

Date of mailing of the international search report

06.02.01

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2260 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Schwachtgen, J-L

INTERNATIONAL SEARCH REPORT

Interr	al Application No
PCT/EP 00/10430	

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 540 065 A (SALK INST FOR BIOLOGICAL STUDI) 5 May 1993 (1993-05-05) figures 4,11; examples ---	1-3,6,9, 11,12, 14,16, 17,23, 24, 26-31, 39,47, 49-54, 57,62, 65,67-73
Y	WO 98 54311 A (BARBAS CARLOS F III ; GOTTESFELD JOEL M (US); SCRIPPS RESEARCH INST) 3 December 1998 (1998-12-03) the whole document ----	1-21,23, 24,26-73
Y	BURCIN M M ET AL: "A REGULATORY SYSTEM FOR TARGET GENE EXPRESSION" FRONTIERS IN BIOSCIENCE, vol. 3, 1 March 1998 (1998-03-01), pages 1-7, XP000874619 the whole document -----	1-21,23, 24,26-73

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP 00/10430

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 47–68 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. Claims Nos.: 22 25
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210
3. Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.
- No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.2

Claims Nos.: 22 25

Present claims 22 and 25 relate to fusion proteins which are defined by the nucleic acid sequence SEQ ID Nos 1-18 of the expression plasmids comprising the coding sequence of said fusion protein. Such a definition does not allow a meaningful database search to be carried out as it is not possible to ascertain from either the figures or the description of the present application where the coding sequence of said fusion proteins is located within the expression plasmid.

The claims 22 and 25 are thus considered to lack clarity within the meaning of Article 6 PCT to such an extent as to render a meaningful search of the claims impossible.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Intern	al Application No
PCT/EP 00/10430	

Patent document cited in search report		Publication date		Patent family member(s)		Publication date
WO 9818925	A	07-05-1998		AU 6908998 A EP 0935657 A		22-05-1998 18-08-1999
EP 0540065	A	05-05-1993		US 4981784 A AT 124721 T AT 182685 T AU 628312 B AU 2818889 A AU 665039 B AU 3026892 A DE 3854120 D DE 3856354 D DE 3856354 T DK 136890 A EP 0325849 A ES 2073408 T IE 68590 B JP 10295385 A JP 10279599 A JP 3006716 B JP 3503597 T KR 9709951 B WO 8905355 A US 5571692 A US 5548063 A US 5171671 A US 5599904 A US 5274077 A		01-01-1991 15-07-1995 15-08-1999 17-09-1992 05-07-1989 14-12-1995 22-04-1993 10-08-1995 02-09-1999 16-12-1999 01-06-1990 02-08-1989 16-08-1995 26-06-1996 10-11-1998 20-10-1998 07-02-2000 15-08-1991 19-06-1997 15-06-1989 05-11-1996 20-08-1996 15-12-1992 04-02-1997 28-12-1993
WO 9854311	A	03-12-1998		US 6140466 A AU 7800398 A EP 0988377 A		31-10-2000 30-12-1998 29-03-2000